. 507
1961

Marks

ol. 75
No. 3

Psychological Monographs

General and Applied

An Assessment of the Diagnostic Process in a Child Guidance Setting

By

Philip A. Marks

Kansas University Medical Center

Price $2.00

Edited by Norman L. Munn

Published by the American Psychological Association, Inc.

Psychological Monographs:

General and Applied

Combining the *Applied Psychology Monographs* and the *Archives of Psychology* with the *Psychological Monographs*

NORMAN L. MUNN, Editor

Department of Psychology, Bowdoin College
Brunswick, Maine

Consulting Editors

ANNE ANASTASI
FRANK A. BEACH
ARNOLD M. BINDER
W. J. BROGDEN
ROBERT R. BUSH
JOHN F. DASHIELL
JAMES J. GIBSON
D. O. HEBB
EDNA HEIDBREDER
JOHN E. HORROCKS
FRANCIS W. IRWIN
JAMES J. JENKINS
BOYD MCCANDLESS
DONALD W. MACKINNON
QUINN MCNEMAR
LORRIN A. RIGGS
CARL R. ROGERS
DAVID SHAKOW
RICHARD L. SOLOMON
ROSS STAGNER

Manuscripts and correspondence on editorial matters should be sent to the Editor. *Psychological Monographs* publishes comprehensive experimental investigations and other psychological studies which do not lend themselves to adequate presentation as journal articles. Major space is given to the author's original contribution; introductory and bibliographic materials, as well as statistical tables and graphs, must be kept within reasonable bounds. Tables, graphs, and appendix materials which deal with detail not essential to adequate presentation of the findings may be made available through the American Documentation Institute—for details of this procedure, see the APA *Publication Manual*. Preparation of manuscripts for publication as monographs should follow the procedure given in the APA *Publication Manual*. Publication in *Psychological Monographs* is free of cost to the author, except in cases where early publication is requested or author's alterations are made in galley proofs.

ARTHUR C. HOFFMAN
Managing Editor

HELEN ORR
Promotion Manager

Correspondence on business matters should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St., N.W., Washington 6, D.C. Address changes must arrive by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee third-class forwarding postage.

COPYRIGHT, 1961, BY THE AMERICAN PSYCHOLOGICAL ASSOCIATION, INC.

Vol. 75, No. 3 Whole No. 507, 1961

Psychological Monographs: General and Applied

AN ASSESSMENT OF THE DIAGNOSTIC PROCESS IN A CHILD GUIDANCE SETTING[1]

PHILIP A. MARKS[2]

Kansas University Medical Center

THERE exists in the field of child psychiatry today a general conviction that unless the milieu of a relationship that is central to a child can undergo favorable change, the most skillful therapeutic intervention is of little avail. This conviction has directed the interests of many clinicians toward investigating the dynamics of primary group interaction and toward conceptualizing the pathology of the child within the psychosocial structure of the family (Ackerman, 1958; Ackerman & Behrens, 1956; Hollis, 1954). Among a majority of child guidance practitioners, it is axiomatic that psychodiagnosis[3] has become the *sine qua non* for planning effective treatment.

Considering the efforts recently extended in assessing the validity of currently popular diagnostic techniques, it is surprising that little, if any, attention has been directed toward evaluating the diagnostic process itself. It is well known, for example, that in the practical situation psychologists typically employ several kinds of psychometric data together with social history, interview, and observation, rather than relying solely upon the blind analysis of a single source of information. Yet, the efficacy of this traditional, time consuming approach, is largely unknown.

The present study was designed to investigate personality characteristics of parents

---

[1] This paper represents a portion of a doctoral dissertation submitted in partial fulfillment of the requirements for the PhD degree at the University of Minnesota, 1959.

The study was made possible by the collaboration of the entire staff of the Amherst H. Wilder Child Guidance Clinic in St. Paul, Minnesota. The opinions and conclusions expressed are those of the writer and do not necessarily reflect the views or endorsement of this clinic.

I wish to express my appreciation to Hyman S. Lippman, clinic director, and to Ethel Harrison, program analyst of the Amherst H. Wilder Foundation, for their full cooperation and support in making clinic facilities available. I am especially grateful to Marilyn Lee, coordinator of case service, whose generous assistance and participation were essential contributions to this study.

The following clinic psychiatrists, psychiatric residents, clinical psychologists, psychiatric caseworkers, and psychiatric group workers gave generously of their time in compiling the numerous *Q* sorts, and I am deeply grateful: Irving I. Gottesman, Christina Herness, Reuben T. Jessop, Richard C. Kogl, Hyman S. Lippman, Karl Oberdieck, Walter M. Ortman, Murray K. Reed, Sylvia D. Reisman, Rene Schwartz, Beth Slocum, Ivan W. Sletten, David S. Thorsen, Grace Weyker, Ester Yanai.

In addition to the clinic staff, I am indebted to the seven psychologists who served as judges and who performed the formidable tasks set forth in the design. To Mildred L. Byrum, Harold Gilberstadt, Sarah F. Holbrook, Wentworth J. Quast, Charles E. Werts, Robert D. Wirt, and Lawrence A. Young, I am especially grateful.

Beyond credit given in later pages, I owe much to a number of persons who contributed not only their time and skills but also their expressions of interest and encouragement. Among these are Paul E. Meehl, who provided *Q* items, reviewed the research design, and made valuable suggestions and recommendations; Walter B. Studdiford, who made materials available from the Ford Foundation Research Project on Diagnosis in Psychiatry; Peter F. Briggs, who provided Minnesota normative data for the MMPI; William Seeman, who helped edit the revised manuscript; Sandra L. Marks, who gave generously her time and support; Robert D. Wirt, my advisor and mentor—it is a privilege to be his first *student*.

Finally, I wish to thank the 48 families who represent the subjects of this study and without whose participation the study would never have been possible.

[2] Formerly at the University of Minnesota.

[3] Throughout this paper, when the term diagnosis is used, personality assessment is implied.

of children referred to a clinic for psychiatric treatment, and to assess the relative efficiency of various sources of data considered by child guidance practitioners in arriving at decisions.

The questions investigated were as follows:

1. Are there personality variables which distinguish parents of children referred to clinics for psychiatric treatment from adults of the general nonpsychiatric population?

2. Is the diagnostic interview, as employed in child guidance clinics, the most efficient source of information about mothers of clinic patients? For example, how accurate are personality descriptions of mothers derived by judges from blind readings of the MMPI?

3. How efficiently do clinical psychologists use their diagnostic time? Are the tests and techniques employed by them the best suited for providing information about clinic patients? For example, how accurate are personality descriptions of *children* derived by judges from blind readings of the MMPI profiles of *parents?*

4. Can clinicians derive personality descriptions of mothers and children which will better predict a criterion than stereotype descriptions derived by the same clinicians, and composite stereotype descriptions derived by other clinicians?

Hypotheses

Five major hypotheses were formulated to answer the questions stated above.

Hypothesis I. There are personality characteristics which distinguish mothers of children referred to clinics for psychiatric treatment from adult females of the general nonpsychiatric population.

Hypothesis II. There are personality characteristics which distinguish fathers of children referred to clinics for psychiatric treatment from adult males of the general nonpsychiatric population.

Hypothesis III. From blind readings of MMPI profiles of mothers, judges can derive personality descriptions of mothers of child guidance patients which will correlate more highly with criterion descriptions derived by clinic caseworkers than with diagnostic descriptions derived by the same caseworkers.

a. Judges' MMPI-based descriptions of mothers will correlate more highly with a criterion than will mother stereotype descriptions derived by the same judges.

b. Judges' MMPI-based descriptions of mothers will correlate more highly with a criterion than will mother stereotype descriptions derived by clinic caseworkers.

c. Caseworkers' diagnostic descriptions of mothers will correlate more highly with a criterion (subsequently derived by the same caseworkers) than will mother stereotype descriptions derived by the same caseworkers.

d. Caseworkers' diagnostic descriptions of mothers will correlate more highly with a criterion (subsequently derived by the same caseworkers) than will composite mother stereotype descriptions derived by other clinic caseworkers.

Hypothesis IV. From blind readings of MMPI profiles of *parents*, judges can derive personality descriptions of *children* referred to psychiatric clinics for treatment which will correlate more highly with criterion descriptions derived by clinic therapists than with diagnostic descriptions derived by clinic psychologists.

a. Judges' descriptions of children, derived from parent MMPI profile data, will correlate more highly with a criterion than will child stereotype descriptions derived by the same judges.

b. Judges' descriptions of children, derived from parent MMPI profile data, will correlate more highly with a criterion than will composite child stereotype descriptions derived by clinic psychologists.

Hypothesis V. From psychological examination data, psychologists can derive personality descriptions of children referred to psychiatric clinics for treatment which will correlate more highly with a criterion than will descriptions of the same children

derived by judges from blind readings of the MMPI profiles of parents.

a. Psychologists' diagnostic descriptions of children will correlate more highly with a criterion than will child stereotype descriptions derived by the same psychologists.

b. Psychologists' diagnostic descriptions of children will correlate more highly with a criterion than will composite child stereotype descriptions derived by other clinic psychologists.

Procedure

The procedure in the present study was designed to obviate a major criticism often leveled against clinical research in a service setting: namely, that in such research, the investigator, under the guise of expediency, often contrives spurious methodological procedure. He promotes various changes in established routine which may implement his purpose and fulfill the requirements of rigorous experimental design, but which are rarely to be duplicated once his data are in. While this criticism does not apply to all research in such settings, it does apply to most research purported to have pragmatic implications. Two recent investigations designed to assess the validity of diagnostic instruments can be taken as cases in point (Little & Shneidman, 1959; Silverman, 1959). The present study, however, necessitates only minimal departure from the reality of "what is done" in a service setting, while hopefully not succumbing to the pitfall of weak or inadequate experimental design.

Selection of Cases

Forty-eight *clinic* cases (i.e., 48 mothers and 48 children), representing 90% of consecutive referrals accepted for treatment from March 15, 1958 to July 31, 1958 comprised the MMPI normative group. Of these, 42 cases in which the mother and child remained in treatment for a minimum of 5 hours and in which the child was free of gross organic (cerebral) impairment and/or mental deficiency, comprised the validation group. The normative and validation samples were presumably from the same population of cases accepted for treatment. However, owing to the small difference between groups (six cases), it was not deemed feasible to test the assumption statistically. Thus, descriptive data are presented for the total sample of 48, whereas validation findings, by definition, were restricted to the sample of 42.

Children

Descriptive data were obtained from the 48 children in the normative group on the major variables of sex, age, intelligence, referral source, and referral complaint. Where possible, these were compared with similar data from the general clinic population (G. R. Patterson, 1955), and from regional studies of comparable groups (Maas, Kahn, & Sumner, 1955; O'Neal & Robins, 1958; Roach, Gurrslin, & Hunt, 1958; Stevens, 1954).

Seventy-five percent of the children were male. This finding was consistent with data reported by other investigators (Maas et al., 1955; O'Neal & Robins, 1958; Roach et al., 1958) indicating a tendency toward a disproportionate representation of males in child guidance clinic populations. Table 1 compares the sex distribution of the sample with the sex distribution of the general clinic population. Although the proportion of males was slightly greater in the sample, the difference between proportions did not approach an acceptable level of significance. ($p = .70$). With respect to sex, the proportion of males to females selected for treatment was representative of the sex ratio among unselected clinic referrals.

The treatment sample of males averaged 10.4 years of age with a range of 4.0 to 16.9 (median = 10.0; $SD = 3.53$) and the females averaged 9.6 years of age with a range of 4.2 to 16.7 (median = 9.5; $SD = 4.03$). Testing for a difference between means did not yield a significant value ($t = .66$). Tables 2 and 3 present male and female age-grouped distributions for the sample and general clinic population. A chi square analysis did not yield a probability value significantly greater than chance for data in either table. Thus the age-grouped distribution of males and females in the

TABLE 1

Sex Distribution of Treatment Sample and Clinic Population

Sex	Treatment Sample %	Clinic Population[a] %
Male	75.0	69.9
Female	25.0	30.1
Both sexes	100.0	100.0

[a] Clinic population data taken from Minnesota Out-Patient Statistical Report of cases terminated ($N = 206$) by the Amherst H. Wilder Clinic during the fiscal year 1956-57.

TABLE 2

AGE DISTRIBUTION OF MALES OF TREATMENT SAMPLE AND CLINIC POPULATION

Age	Treatment Sample %	Clinic Population[a] %
Under 5	5.7	9.8
5–9	38.8	37.5
10–13	36.1	33.3
14–17	19.4	19.4
All ages	100.0	100.0

Note.—Chi square = 1.01 with 3 *df* (p = .80).
[a] Clinic population data taken from Minnesota Out-Patient Statistical Report of cases terminated (N = 206) by the Amherst H. Wilder Clinic during the fiscal year 1956-57.

sample was representative of similar distributions among unselected clinic cases.

During the diagnostic period, 90% of the children were administered either the Stanford-Binet or the Wechsler Intelligence Scale for Children. The average IQ for 33 males was 100.7 with a range of 71 to 143 (median = 100.0; SD = 16.59). The average IQ for 10 females was 102.0 with a range of 81 to 139 (median = 102.0; SD = 15.18). As the data suggest, the sexes were similar in IQ distribution. Table 4 presents the results of a comparison between the IQ distribution of the sample and the IQ distribution of children referred to three Twin City[4] child guidance agencies (G. R. Patterson, 1955). For computational purposes, the IQs were distributed among three categories: below average (71-90), average (91-110), and above average (111-145). The average IQ of the sample was 101.0, whereas the average IQ of 213 unselected agency referrals was 105.6.

The findings presented above and in Table 4 offer some observations of general interest. First, the IQ distribution of children selected for treatment closely approximates the theoretical distribution of the normal population. Second, the IQ distribution among unselected clinic referrals is negatively skewed, and the mean of the distribution is numerically higher than the mean of the sample. Unfortunately, the source of these data does not provide enough information to test the difference statistically. The direction of the difference, however, was consistent with findings of other investigators (Roach et al., 1958; Stevens, 1954). Third, the differences between the grouped distributions, while not of acceptable statistical significance (p = .10), suggest that a disproportionately large number of children with below average intelligence estimates were selected for treatment. Although this finding is contrary to expectations, it is nevertheless understandable in view of the major source of clinic referrals and the most frequently offered referral complaint. As will be reported below, school referrals accounted for over 40% of children selected for treatment and, in 45% of the treatment cases, "inability to achieve in school" was listed as a presenting complaint. A general conclusion which emerged from these data was that school exigencies, to a large extent, determined which children would be selected for treatment. Without prejudging the complex relationship between academic failure and emotional distress, the data suggest two paradoxical phenomena. The first concerns the question of what, in the interest of economy, is the

[4] Minneapolis and St. Paul.

TABLE 3

AGE DISTRIBUTION OF FEMALES OF TREATMENT SAMPLE AND CLINIC POPULATION

Age	Treatment Sample %	Clinic Population[a] %
Under 5	8.3	16.2
5–9	50.0	43.5
10–13[b]	25.0	22.6
14–17[b]	16.7	17.7
All ages	100.0	100.0

Note.—Chi square = .74 with 2 *df* (p = .70).
[a] Clinic population data taken from Minnesota Out-Patient Statistical Report of cases terminated (N = 206) by the Amherst H. Wilder Clinic during the fiscal year 1956-57.
[b] Groups 10-13 and 14-17 were combined as recommended by Siegel (1956) for cells with small Ns.

TABLE 4

IQ DISTRIBUTION OF TREATMENT SAMPLE AND THREE TWIN CITY CHILD GUIDANCE AGENCIES[a]

IQ	Treatment Sample %	Twin City Agencies %
Below average (71–90)	25.5	17.4
Average (91–110)	58.1	50.2
Above average (111–145)	16.4	32.4
Totals	100.0	100.0

Note.—Chi square = 4.83 with 2 *df* (p = .10).
[a] Data taken from G. R. Patterson (1955). Twin City sample comprises cases from the Amherst H. Wilder Clinic, the Washburn Memorial Clinic, the Child Study Department, and the In-Patient Psychiatric Service, University of Minnesota Hospitals.

most expedient use of psychiatric time. The second involves the known positive correlation between intelligence estimates and academic success. If one holds that psychotherapy (at best) is a learning situation, then children with high intelligence, *ceteris paribus,* would be more likely to respond than children not so endowed. Suffice it to say that intelligence was *not* the prime criterion for selecting the treatment cases in the present sample.

Data pertaining to source of referral are categorized in Table 5. With exception of public and private agency cases, the categories shown represent the major sources reported by other metropolitan clinics (Maas et al., 1955; Stevens, 1954). Since in most instances, public and private agency referrals are made for diagnostic purposes only, and since the sample was restricted to cases entering treatment, this source was omitted and the percentages shown were recomputed from subtotals of the sample distributions. In the absence of data for the general clinic population, samples from regional clinics are presented for comparison. Testing for a difference between groups yielded a significant value ($p = .01$). Thus, the four distributions were dissimilar in their proportion of cases referred by different sources.

Among the treatment sample, school referrals comprised the largest single source of clinic intake, and were proportionately greater than referrals from any other source among all clinics reported. It would seem likely that in larger cities (e.g., Chicago, New York, and San Francisco) where more numerous facilities exist, school personnel would be less inclined to refer problem cases to child guidance practitioners. On the other hand, where fewer facilities exist, patient loads are usually larger and there should be less available time for academic problem cases. It is equally probable that certain clinics choose to focus upon different problems (e.g., mental retardation, organic impairment, schizophrenia). What is apparent, however, is that clinics differ in their orientation and referral policies and that samples taken from any one of them are not necessarily representative of other agencies.

A list of symptoms and referral complaints which characterized the sample at the time of intake appears in Appendix A. The list employed was developed by G. R. Patterson (1955) in an earlier study of Twin City clinic referrals. The incidence ratings reported were based upon descriptions of children's behavior reported by the parents or the referring agency. All tabulations were made by the investigator who read each intake report and recorded every symptom and complaint appearing therein. Table 6 compares the five most frequently reported complaints for children in the sample with their frequency among unselected clinic patients. It can be seen from Table 6 that the groups yielded a difference that was statistically significant beyond the .001 level. Table 7 compares the five most frequently reported complaints for unselected clinic referrals with their frequency among the treatment sample. Again, a significant difference was found ($p = .05$).

Before commenting on these findings, a word of caution is necessary. Although first consideration was given to objectivity of recording (i.e., only symptoms and complaints explicity stated were tabulated) no reliability estimates were available for the ratings. Furthermore, no attempt was made to distinguish "major" from "minor" complaints, and thus the data do not necessarily reflect

TABLE 5

SOURCES OF REFERRAL

Source	Treatment Sample %	Chicago[a] (N = 200) %	New York[b] (N = 347) %	San Francisco[b] (N = 306) %	Total (N = 901) %
School	43.4	22.7	26.1	20.6	24.6
Self, relative, etc.	23.9	34.5	31.8	30.9	31.5
Physician, hospital	23.9	26.2	25.7	38.9	30.3
Courts	8.7	16.5	16.4	9.5	13.5
Totals	99.9	99.9	100.0	99.9	99.9

Note.—Chi square = 27.20 with 9 df ($p = .01$).
[a] Taken from Roach (1958). Based on intake at the Guidance Center of Buffalo.
[b] Taken from Maas et al. (1955). Based on intake at six New York clinics (Jewish Board of Guardians, New Rochelle, Northside, Orange, White Plains, and Yonkers) and three San Francisco clinics (Children's Hospital, Langley Porter, and Mt. Zion).

TABLE 6

INCIDENCE OF FIVE MOST FREQUENT COMPLAINTS AMONG CHILDREN IN THE TREATMENT SAMPLE WITH CORRESPONDING INCIDENCE AMONG THE GENERAL CLINIC POPULATION

Referral Complaint	Treatment Sample %	Clinic Population[a] %
Unable to achieve in school	45.8	48.5
Poor peer relationships	25.0	11.8
Nervous	25.0	8.1
Sleep disturbances	25.0	.8
Fearful	25.0	4.4

Note.—Chi square = 32.20 with 4 df ($p = .001$).
[a] Taken from G. R. Patterson (1955). Data represent an unselected group of 136 cases referred to the Amherst H. Wilder Clinic.

TABLE 7

INCIDENCE OF FIVE MOST FREQUENT COMPLAINTS AMONG CHILDREN IN THE GENERAL CLINIC POPULATION WITH CORRESPONDING INCIDENCE AMONG THE TREATMENT SAMPLE

Referral Complaint	Clinic Population[a] %	Treatment Sample %
Unable to achieve in school	48.5	45.8
Defiant, nonconforming, disobedient	16.9	6.2
Immature	12.5	2.1
Poor peer relationships	11.8	25.0
Unable to concentrate	11.8	12.5

Note.—Chi square = 10.99 with 4 *df* (p = .05).
[a] Taken from G. R. Patterson (1955). Data represent an unselected group of 136 cases referred to the Amherst H. Wilder Clinic.

the major problems for which the children were referred. Most important, it is not known whether the data from the two samples were recorded under similar conditions. It is possible that the difference obtained represented variations in recording procedure rather than a difference in types of complaints presented by children in the groups. Having stated this precaution, certain observations are noteworthy. The most striking finding is the relatively high incidence of school achievement problems in both groups. For approximately half of all clinic referrals, and for a proportionate number of children selected for treatment, "inability to achieve in school" was offered as a presenting complaint. This finding was consistent with data showing a disproportionately large number of children in treatment referred by schools. Thus, where the school served as the major source of clinic referrals, school achievement problems constituted the most frequently offered referral complaint. It can also be seen from Tables 6 and 7 that behavior which is most offensive to the culture ("defiant," "nonconforming," "disobedient," and "immature"), and for which society presumably makes referrals, is relatively infrequent among children selected for treatment. In contrast, behavior presenting the most subjective discomfort ("nervous," "fearful," and "sleep disturbances"), has a high incidence among children selected for treatment, but was rarely present among unselected clinic cases.

Parents

Descriptive data were obtained for age, marital status, and education of the children's parents, and for socioeconomic status, religion, and racial origin, of the children's families. Where possible, these data were compared with norms for the general clinic population, and with Minnesota male and female standardization data (revised) for the MMPI (Hathaway & Briggs, 1957).

The mothers averaged 37.1 years of age and ranged from 27 to 56 (median = 35.0; SD = 7.03), and the fathers averaged 40.7 years of age and ranged from 26 to 55 (median = 38.0; SD = 4.94). In 26% of the families the mother was older than the father (mean = 1.5 years; range 1 to 3). Table 8 compares age data for the sample with age characteristics of the MMPI standardization group. It can be seen from Table 8 that the sample means and variances of both sexes differ significantly. The parent sample was older and more homogeneous than adults representative of the general Minnesota population. This finding, however, was to be expected. The sampling of parents precludes a normal age distribution. A comparison of the age distribution for both groups revealed that 27% of males and 24% of females included in the MMPI standardization group were classified within the adolescent age range (16 to 25 years), whereas none of the parents in the sample were younger than 26 years of age.

Thirty-seven of the 48 families studied were intact (i.e., were not disrupted by death, separation, or divorce) at the time of intake. Of the remaining 11, 7 were broken due to divorce and 4 were disrupted by fathers' deaths.

Educational characteristics of the parent sample and the MMPI standardization group are given in Table 9. The fathers averaged 10.1 years of education and ranged from 8 to 16 (median = 10.0; SD = 2.59), and the mothers averaged 10.3 years of education and ranged from 7 to 17 (median = 10.5; SD = 2.55). Among the MMPI group, males averaged 9.7 years of education and ranged from 3 to 18 (median = 8.0; SD = 2.78), and females averaged 10.5 years of education and ranged from 3 to 18 (median = 10.0; SD = 2.83). Although the sample distribution was negatively skewed, testing for differences between means and variances failed to yield ratios significant at an acceptable level of confidence. With respect to years of education, the parents in the sample were

TABLE 8

AGE CHARACTERISTICS OF PARENT SAMPLE AND MMPI STANDARDIZATION GROUP

Subjects	*N*	Mean	*SD*	*t*	*F*
Males					
Parent	42	40.7	4.94		
MMPI	226	34.3	10.93	6.16**	4.89**
Females					
Parent	48	37.1	7.03	2.92**	2.13*
MMPI	315	34.5	12.02		

* Significant between the .10 and .05 levels.
** Significant at the .01 level or beyond.

TABLE 9

EDUCATIONAL CHARACTERISTICS OF PARENT SAMPLE AND MMPI STANDARDIZATION GROUP

Subjects	*N*	Mean	*SD*	*t*	*F*
Males					
Parent	42	10.1	2.59	.08	1.11
MMPI	226	9.7	2.78		
Females					
Parent	48	10.3	2.55	.04	1.23
MMPI	315	10.5	2.83		

representative of adults in the general Minnesota population.

The Minnesota Scale of Paternal Occupations (Goodenough & Anderson, 1931), was used as an index of family social status. In Table 10, the occupational distribution of the parent sample is compared with distributions from the clinic population and the MMPI standardization group. Farmers (Class IV) have been excluded owing to their low frequency (.3%) in the clinic population and their nonexistence in the other samples. Testing for a difference between groups yielded a value significant at the .001 level. The three sample distributions did not come from the same homogeneous population.

Table 10 shows that the clinic extends diagnostic service to all occupational groups (except Farmers), but with some underrepresentation of the lower class. The fact that 70% of the families were representative of Class III and Class V occupations, suggests that treatment was a service extended primarily to the middle class. This finding was not in keeping with reports of other investigators which show proportionately more upper-class families among child guidance referrals (Maas et al., 1955; Roach et al, 1958), and a proportionately greater number of upper-class families receiving treatment services (Hollingshead & Redlich, 1958). There are undoubtedly many interrelated factors which determine a given clinic's case load. Referral source, intake policy, and general orientation toward treatment have all been mentioned above. It is quite possible that in clinics where medical referrals are proportionately few, and where school referrals comprise the largest single source of intake, the social class representation might be expected to approximate the normal distribution.

TABLE 10

OCCUPATIONAL CHARACTERISTICS OF PARENT SAMPLE, CLINIC POPULATION, AND MMPI STANDARDIZATION GROUP

Occupational Groups	Parent Sample %	Clinic Population[a] %	MMPI Group %
I Professional	8.3	5.1	7.8
II Semi-professional	4.2	14.9	5.2
III Skilled, clerical, business	39.5	15.6	19.4
V Semi-skilled, minor clerical	31.2	32.2	23.0
VI Slightly-skilled	12.5	18.0	35.9
VII Day laborers	4.2	14.2	8.6

Note.—Chi square = 75.03 with 10 *df* ($p = .001$).
[a] Taken from G. R. Patterson (1955). Clinic population data represent a combined distribution of homogeneous samples from three Twin City agencies.

No data were available pertaining to the religious composition of the clinic population. Among the families in the treatment sample, 54% were Protestant, 29% were Catholic, 5% were Jewish, and 12% expressed no preference or their denomination was unknown.

Only one family in the normative group was nonwhite as compared with an estimated 5% nonwhite in the clinic population. Since this case terminated before reaching the criterion number of treatment interviews, 100% of the validation sample were Caucasian.

Clinical Evaluators

Child Psychologists

Two staff psychologists and two third-year clinical psychologist trainees served as child diagnosticians. They differed markedly in academic training, theoretical orientation, years of experience, clinical interests, and bias toward various tests. In Table 11, the psychologists are identified

TABLE 11

DESCRIPTIVE DATA CONCERNING CHILD PSYCHOLOGISTS

Psychologist	Months of Clinical Experience			Diagnostic *Q* Sorts Children ($N = 42$)	
	Study Clinic	Elsewhere	Total	*N*	%
A	120	120	240	23	54.8
B	54	24	78	11	26.2
C[a]	22	6	28	4	9.5
D[a]	12	6	18	4	9.5

[a] Third-year clinical psychologist trainee.

as A, B, C, and D, and descriptive data are presented for each.

At the time of intake, each family was assigned to one of four clinic teams and all diagnostic appointments, in addition to the date of the diagnostic conference, were scheduled in advance of the family's first formal clinic contact. The diagnostic period (i.e., the time between intake and the diagnostic conference) ranged from 13 to 46 days and averaged 27 days. The number of children assigned to each psychologist varied greatly and was ultimately determined by such factors as size of case load, preference, and training expediencies. Consequently, trainees, who were employed on a half-time basis only, were assigned fewer children than clinic staff. It can be seen from Table 11 that over one-half the children were assigned to Psychologist A, whereas only one-tenth were assigned to Psychologists C and D. Moreover, the two staff psychologists evaluated 81% of the children as compared with 19% evaluated by the two psychologist trainees. The diagnostic evaluations of children were essentially a staff enterprise.

A description of the tests and techniques used by the psychologists is of major importance, especially since the accuracy of judgments derived from them was the focus of concern. It has been mentioned above that a major consideration of the design was to adhere to established clinic routine throughout the study. Accordingly, psychologists were permitted to select whatever instruments they desired, or to spend their time however they chose. Table 12 presents a list of the tests employed and their frequency of use. The data in Table 12 reveal several interesting facts about the clinical interests and test biases of the clinic psychologists. First, the Draw-A-Person technique, despite recurrent reports of invalidity (e.g., Swensen, 1957), was the most popular instrument in the psychologists' armamentarium. Perhaps this can be justified in part by its professed use as a rapport-gaining device, and in part, by its use as an estimate of intelligence. Second, infrequently used were various projective techniques (e.g., TAT, CAT, Blacky pictures) which are presumably unique in their contribution of dynamic content. Third, achievement and intelligence tests were commonly used and, with few exceptions, were found to complement the Rorschach and Draw-A-Person in forming a basic battery.

TABLE 12

PSYCHOLOGICAL TESTS AND TECHNIQUES ADMINISTERED TO CHILDREN
(N = 39)[a]

Instrument	N	%
Draw-A-Person	34	87.1
Rorschach	30	76.9
Reading Achievement[b]	29	74.3
Stanford Binet (Forms L or M)	28	71.7
Arithmetic Achievement[b]	20	51.2
Spelling Achievement[b]	16	41.0
WISC	12	30.7
Sentence Completion	11	28.2
Bender-Gestalt	9	23.1
Make-A-World	8	20.5
Fables, 3 Wishes, etc.	4	10.2
TAT	4	10.2
Blacky Pictures	4	10.2
CAT	3	7.7
Ammons Picture Vocabulary	3	7.7

[a] Excluding three children who refused testing.
[b] Wide-Range Achievement Test.

The tests were also grouped into the major areas of personality, intelligence, and achievement. Table 13 gives the source and amount of data available to clinic psychologists at the time they made their diagnostic judgments. It can be seen from Table 13 that the modal psychological examination of children consisted of an intelligence estimate, three personality measures, and in most instances, some index of academic progress. For all but three children, the examinations seemed optimal since they covered three major areas of function in addition to providing opportunity for observation and interview. Evaluating data for the three children who refused testing were obtained from observation and interview.

It is important to emphasize that the relative accuracy of psychologists' *descriptive and inferential statements* about children, derived from an integration of psychometric, interview, and observational data (i.e., the clinician-as-an-instrument), was the variable under investigation. The diagnostic process was a matter of concern only in the degree to which the *psychologist* contributed to it.

Following completion of the psychological examination and prior to the diagnostic conference, the psychologist compiled a Q-sort personality description of the child. Each Q-sort description constituted a structured and comprehensive psychological report and served as the basis for a comparative evaluation with similar descriptions independently derived from the MMPI.

TABLE 13

SOURCE AND AMOUNT OF DATA AVAILABLE TO PSYCHOLOGISTS FOR MAKING CLINICAL JUDGMENTS
(N = 39)[a]

Source	N	%
Personality[b]	39	100.0
Intelligence	38	97.4
Achievement	32	82.0

[a] Excluding three children who refused testing.
[b] Mean = 2.67 instruments per child.

Reliability data for child diagnostic descriptions were obtained from each clinic psychologist. The two psychologist trainees repeated sorts of their final case 2 days following their original *Q*-sort descriptions. Reliability estimates were obtained from each staff psychologists for each of five children assigned to them for treatment. The latter resort descriptions also served as criterion measures and were completed approximately 3 months following the date of the diagnostic *Q*-sort descriptions.

Social Caseworkers

Diagnostic and criterion *Q*-sort descriptions of mothers were compiled by four staff psychiatric caseworkers. Each had been trained at the University of Minnesota School of Social Work, and had had at least 7 years of casework experience. Their experience at the study clinic ranged from 2 to 12 years. In Table 14, the caseworkers are identified as W, X, Y, and Z, and descriptive data are presented for each.

The caseworkers' initial task involved soliciting parent cooperation in the testing program. Parents were first apprised of the testing request by the coordinator of casework service (Chief Social Worker) at the time of intake. They were told that the clinic was engaged in a research program designed ultimately to aid in a better understanding of parent-child relations and that all parents were requested to complete a personality inventory (MMPI) early in the course of diagnostic study. Parents were then reminded of the testing request at the close of their first diagnostic appointment. Caseworkers were responsible for scheduling time sufficient to insure completion of the inventory before the date of the diagnostic conference. In the majority of cases, appointments were scheduled the hour preceding or following the second diagnostic interview. The clinic receptionist administered tests and collected the booklets and answers sheets. The latter were removed from the clinic daily by the investigator. No professional personnel had access to MMPI data at any time during the course of the study proper. Test data were made available to them only upon completion of criterion *Q*-sort descriptions. As a result of the above procedure, every mother complied with the testing request. The return for fathers was 65% ($N = 32$). It is noteworthy that no excessive pressure was extended to secure parent cooperation and, in retrospect, it was believed that had greater effort been made, more fathers would have cooperated.

TABLE 14

DESCRIPTIVE DATA CONCERNING PSYCHIATRIC CASEWORKERS

Caseworker	Months of Casework Experience			*Q* Sorts[a] Mothers ($N = 42$)	
	Study Clinic	Elsewhere	Total	*N*	%
W	145	72	217	15	35.7
X	98	0	98	11	26.2
Y	87	9	96	10	23.8
Z	25	60	85	6	14.3

[a] Includes both diagnostic and criterion descriptions.

The diagnostic period for parents was identical in length to that for children (cf. children above). The number of scheduled interviews during this period ranged from 2 to 5 and averaged 3. The number of parents assigned to each caseworker was primarily determined by availability and case load.

Upon completion of the final diagnostic interview and prior to the diagnostic conference, each caseworker compiled a *Q*-sort description of the mother using the *same items* employed by the diagnostician and therapist in describing the child. Each description constituted a diagnostic report and served as the basis for comparative evaluations with similar descriptions independently derived from the MMPI.

Treatment of the mother was in all instances undertaken by the caseworker assigned diagnostic responsibility for the case. Thus each caseworker who compiled a diagnostic description also compiled a criterion description of the same mother following an average of seven postconference interviews. The number of treatment interviews ranged from three to nine, and the minimum number of total interviews (i.e., diagnostic and treatment combined) was six. The families participated in the study for an average of 104 days (range 56 to 167).

MMPI Judges

Seven nonclinic judges performed the somewhat formidable tasks set forth in the design. There were discernible differences among them in academic status, time spent in clinical settings, type of a setting in which they worked, and experience with the MMPI. Six were clinical psychologists who either had experience in child guidance settings or were familiar with child guidance populations. One was a child psychologist with no clinical experience. Four judges held PhD degrees and three were third-year graduate students, then enrolled in doctoral programs at the University of Minnesota. Finally, in terms of MMPI experience, four judges could be classified as "experts" and three could be classified as "neophytes."

All judges were preselected by the investigator. Two factors were considered of utmost importance in their selection. First, there is mounting evidence which shows that *clinical* predictions,

from adult psychiatric patient MMPI profiles by experienced clinicians may have high external validity, e.g., agreement with therapists (Duker, 1958; Halbower, 1955; Little & Shneidman, 1954), but that predictions by clinicians lacking experience may be only of moderate accuracy (Duker, 1958), or of limited value (Sines, 1957). With few exceptions, the MMPI is *not* the instrument of choice among child guidance practitioners and it is indeed doubtful whether many MMPI experts exist in child guidance settings. Thus, to evaluate the usefulness of the instrument in such settings, it was considered important to estimate the accuracy with which relatively inexperienced persons could predict the criterion, and to determine the accuracy which might be expected under more ideal conditions.[5] Second, another major finding of recent research suggests that a thorough knowledge of the population base rates (i.e., the relative incidence of various behavior patterns among the class of persons under consideration) is essential for the effective use of clinical data (Halbower, 1955; Meehl & Rosen, 1955; Sines, 1957). Thus six of the seven judges were selected because of their familiarity with child guidance populations. Table 15 identifies the judges and reports descriptive data pertaining to each.

Two separate prediction tests were required of judges by the design. First, *Q*-sort descriptions of mothers were compiled from blind readings of MMPI profiles, and second, *Q*-sort descriptions of children were *extrapolated* from parents' MMPI profile data. For both tasks, judges were provided with standard MMPI profiles plus information pertaining to the child's age, sex, grade, and number of siblings; and the parent's age, education, and occupation (hereafter referred to as "minimal data").

Q-sort descriptions of 42 mothers and 42 children were compiled by five of the seven judges. Each judge was provided with both parents' MMPI profiles (when available) and a mimeographed sheet of instructions containing the following minimal data:

> Enclosed is/are the MMPI profile/s of a mother and father whose _____ year old boy/girl in the _____ grade was recently referred to a child guidance clinic. The child is living with both parents/the mother only, _____ brothers and _____ sisters. Both mother and child are currently being seen in treatment. Using the 135 *Q* statements, describe the mother and child as they would look to a therapist knowing them well.

In the absence of profile data for father, the appropriate reason was given (i.e., divorced, deceased, separated, or refused testing). Typically, though not necessarily, the procedure involved compiling mother *Q*-sort descriptions first and child descriptions second. Following each sort the items were recorded and the *Q* deck reshuffled. After completing sorts for all cases assigned, each judge resorted one case selected at random by the investigator from among those previously described. The time lapse between sorts averaged 2 weeks (range 7 to 20 days). The repeat sorts served as the basis for reliability estimates of judges' predictions of mothers and children.

Criterion for Child Descriptions

Seven members of the clinic staff and two residents in psychiatry served as child therapists. They represented each of the three disciplines in the clinic and, like the diagnosticians, differed markedly in academic training and years of clinical experience. The therapists are identified in Table 16 and descriptive data are presented concerning the professional status, clinical experience, and number of children evaluated by each. Table 16 shows that each therapist had at least 3 years of total clinical experience and, with the exception of the two residents and of one psychiatric groupworker, each had spent a minimum of 5 years at the study clinic. The median months of total experience was 90 (range 36 to 444).

Each child-therapist assignment was made at the time of the diagnostic conference. While these were based largely upon the child's needs, staff availability, interest, and training requisites were known to have affected some assignments. Accordingly, 8 children were assigned to activity groups only, whereas 34 children were assigned to therapists for individual treatment. Each therapist knew that his patient was involved in the study and would be rated (*Q* sorted) later in therapy. Among the 34 children selected for individual treatment, 28 (82%) were assigned to psychia-

[5] Findings which pertain to this evaluation are reported in the original paper only (Marks, 1959b).

TABLE 15

DESCRIPTIVE DATA CONCERNING MMPI JUDGES

Judges[a]	Clinical Experience	MMPI-Based *Q* Sorts Mothers and Children (N = 42)[b]	
	Months	N	%
I[c]	168	8	19.0
II[c]	120	9	21.5
IV	108	8	19.0
V	36	9	21.5
VI	12	8	19.0

[a] Excluding Judges III and VII who derived *Q*-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] Read: 42 mothers and 42 children.
[c] MMPI expert.

TABLE 16

DESCRIPTIVE DATA CONCERNING CHILD THERAPISTS

Therapist	Months of Clinical Experience			Criterion *Q* Sorts Children ($N = 42$)	
	Study Clinic	Elsewhere	Total	*N*	%
1	324	120	444	6	14.3
2	120	120	240	3	7.1
3	60	36	96	2	4.8
4	87	9	96	1	2.4
5	66	24	90	9	21.4
6	72	0	72	4	9.6
7[a]	4	54	58	8	19.0
8[a]	5	31	36	5	11.9
9	12	24	36	4	9.5

[a] Resident in child psychiatry.

trists, 5 (15%) were assigned to psychologists, and 1 child was assigned to a caseworker. Thus, for the majority of the cases, where staff psychologists served as child diagnosticians, staff and resident psychiatrists served as child therapists.

After the child had been seen for an average of 10 interviews, the therapist compiled a *Q* sort using the same items employed earlier by the psychologist. The actual number of interviews varied between 5 and 13. Adopting 5 interviews as a minimum was not altogether an arbitrary decision. Meehl[6] has found that correlating *Q*-sort descriptions derived from successive interviews with a criterion of 24 interviews, yields a rapidly accelerating curve which reaches an asymptote between the fourth and fifth interviews. *Ergo,* the decision was made to include all children in the sample who remained in treatment for a minimum of 5 hours, but that therapists would not otherwise compile criterion descriptions until after the tenth interview.

The express purpose of the present design precluded uncontaminated clinic *Q*-sort descriptions. No attempt was made to deprive therapists of psychometric data about children, for such data are normally available to them and therapist's judgments are typically founded to some extent upon data available in psychological reports. Moreover, 5 of the 42 children were assigned for treatment to the psychologist who had evaluated (and *Q* sorted) them diagnostically. The net result was that in 37 cases the criterion descriptions were confounded to some extent by psychometric data, and in 5 cases the diagnostician served as the child therapist.

Notwithstanding the accessibility of clinic folders, test results, and psychological reports, therapists were instructed to rely upon their own therapy notes in deriving criterion descriptions. The latter was intended to effect more accurate descriptions rather than delimit data which were normally available to them.

The major consequence of the above procedure warrants comment. Comparing judgments made by different persons on essentially the same data (or by the same persons on the same data) yields information concerning the reliability of judgment only. However, once such information is obtained, it can usefully serve as an estimate of the maximum validity for judgments independently derived from other sources. For example, one objective of the study was to evaluate the comparative accuracy of psychologists' diagnostic descriptions of children and of descriptions of the same children independently derived by judges from blind readings of parent MMPI profiles. Confounding the criteria with psychometric data would only result in spuriously high estimates of accuracy for clinic psychologists. Thus, under these circumstances, the extent to which judges' descriptions were found to approximate (or to surpass!) the accuracy achieved by psychologists, *would only be enhanced* if the criteria were derived independent of the data which served as the source of the latters' initial judgments. It follows that the criteria must also be derived independent of the MMPI. This was assured of course by removing the answer sheets from the clinic immediately upon completion of testing. MMPI data were inaccessible to therapists prior to their criterion evaluations.

Reliability data were obtained for child criterion descriptions by having five therapists compile resorts of their final case within 2 weeks of their initial criterion descriptions.

Estimating the Accuracy of the Criterion

Owing to the fact that each child criterion description was compiled by a single clinician (the child's therapist), a source of variance in the criterion was introduced which in part can be attributed to individual differences in clinical frames of reference, and in part reflects differences in clinical ability among therapists. That this variance could have deleterious effect upon diagnostic predictions seems obvious; i.e., the most direct effect would be to lower the *Q* correlations (attenuating the true variance) of the various data-based predictions. To obviate such a state of affairs, a precaution was taken to insure at least moderate validity of the criterion. The correlations of therapists' criterion descriptions with psychologists' diagnostic descriptions were compared for each child. Since the criterion was contaminated by psycho-

[6] Personal communication, Paul E. Meehl, November 1958.

metric data, it was decided that the lowest average (acceptable) agreement between clinic therapists and psychologists should exceed chance. Therefore, a therapist whose average correlation across psychologists was not significantly greater than zero at the .001 level or beyond should be eliminated and his descriptions discarded from further analysis. Table 17 presents the average correlations of clinic therapists with clinic psychologists for the 37 children described by both.

Obvious from Table 15 is the fact that *none* of the criterion descriptions compiled by Therapist 9 showed better than chance agreement with diagnostic descriptions compiled by clinic psychologists. This was a most disconcerting finding and suggested a definite lack of communication between this therapist and other members of the clinic staff. For the purpose stated above, the four criterion descriptions compiled by Therapist 9 were discarded and the number of children included in the sample was reduced to 38.

When therapists' total clinical experience (*q.v.* Table 16) was compared with their average agreement with psychologists, and a rank order correlation computed, a coefficient of .57 was obtained ($p = .10$). While not of acceptable statistical significance, it suggested some relationship between diagnostician-therapist agreement and therapist clinical experience.

Selecting the Q Array

A decided advantage of the *Q* sort in psychiatric research is that it facilitates development of a standard language of behavior whereby different clinicians, making independent interpretations of similar or dissimilar data, can communicate impressions in an objective form suitable for statistical evaluation. Furthermore, if the language is so developed as to be suitable for children and adults of both sexes, parent-child comparisons can be made. While the latter was not a feature of the present study, items were selected with this expediency in mind. In addition to age-sex nonspecificity, several other criteria were used to assure the pertinence, ratability, variability, and comprehensive coverage of the statement pool.

TABLE 17

MEAN CORRELATIONS OF THERAPISTS' CRITERION DESCRIPTIONS WITH PSYCHOLOGISTS' DIAGNOSTIC DESCRIPTIONS ($N = 37$)[a]

Therapist[b]	*N*	Mean	Range
1	6	.41	.28—.60
4	1	.47	—
5	9	.29	.05—.48
6	4	.49	.21—.65
7[c]	8	.48	.26—.67
8[c]	5	.27	.20—.44
9	4	−.09	−.33—.16

[a] Excluding five children assigned to clinic psychologists for treatment.
[b] Excluding Therapists 2 and 3 (Psychologists A and B) who compiled diagnostic *Q*-sort descriptions.
[c] Resident in child psychiatry.

Sources of Items

In attempt to secure a comprehensive sample of items representative of the extensive domain of personality, and sufficient for the description of any case, normal or pathological, several sources were explored. Approximately 100 items were obtained by the investigator, who culled descriptive and inferential statements from the folders of cases filed at the study clinic; 295 items were borrowed from a pool compiled earlier by Paul E. Meehl; and 1,538 items were acquired from a pool then under development at the University of Minnesota by the Ford Foundation Research Project on Diagnosis in Psychiatry.

The purpose of the Ford project has been described (Duker, 1958) as "a research program for investigation of personality dimensions having maximum descriptive precision and predictive power in the understanding of psychiatric patients" (p. 64). The initial phase of this program was devoted to the compilation of 6,000 descriptive items which were presumably representative of the personality sphere "entirely and accurately." The following areas were sampled: Interests, avocational; Interests, vocational; Value-orientation: Primary-group relationships; Attitudes; Mood and temperament; Vocational activity; Manifest interpersonal patterns; Psychopathology; Ethical behavior; Self-concept; and Ego-organization and "character structure" (Studdiford, unpublished).

Once compiled, the pool was submitted to "armchair" screening and items were eliminated which were judged to be any of the following: ambiguous, rare, genotypic, unknown to therapists, atomistic, synonymous, extreme, unfamiliar, evaluative, broad, "double-barreled." The intact pool at the time of the present study contained 1,538 items which had survived the stringent screening criteria. As alluded to above, the pool contained *phenotypic* items only, i.e., items which refer to "relatively surface, objective, descriptive, 'summarizing' aspects of behavior" (i.e., manifest traits) as opposed to *genotypic*, which refer to "internal events, states, structures inferred from the phenotype, and often said to 'explain' it" (i.e., latent traits) (Meehl, unpublished). Since the purpose of the present study was to standardize a pool containing both genotypic and phenotypic item content, Meehl's pool served as the primary source of genotypic items. Thus, from the sources mentioned above, approximately 2,000 items were obtained for potential selection.

In view of the relatively objective criteria of age and sex nonspecificity, the pool was first screened by the investigator who eliminated 1,381 items which were either duplicated, or a priori seemed inappropriate for adults and children of both sexes. The remaining 619 items comprised the basic pool, which was then submitted to empirical screening by the study clinic staff.

Screening Items for Clinical Pertinence

The 619 items were then submitted to two staff psychiatrists who were instructed to rate for "clinical pertinence." For an item to be considered pertinent, it had to be nonspecific with respect to age and sex, and had to contain information commonly discussed at team meetings, case conferences, etc., and/or have diagnostic or treatment implications. Although the first criterion was considered by the investigator in the original selection of items, it was believed necessary to have the pool rerated on this criterion to insure elimination of any item missed during the initial screening process. Thus, the judges were instructed first, to evaluate the appropriateness of each item for adults and children, and second, to consider the pertinence of each item for the clinic population. The categories of "suitable," "questionable," and "unsuitable" were designated for this purpose. Any item considered inappropriate for adults and children was automatically rated "unsuitable." All other items were considered pertinent and rated either "questionable" or "suitable."

Of the 619 items submitted for screening, 476 were rated suitable, and 32 were rated questionable by both judges. The interjudge agreement was 82%. For all remaining items the judges were mixed in agreement. Only those items were retained for further screening which either both judges agreed were suitable (476 items), or one judge rated suitable and the other rated questionable. The latter combinations comprised 43 and 56 items, respectively. Of the 619 items thus screened, 575 met the criterion of clinical pertinence.

Screening Items for Ratability

Next, the 575 items were submitted for judgments of "ratability" by each remaining member of the clinic staff. Accordingly, six caseworkers, four psychologists, and three group workers were instructed to judge each item in terms of its ratability based upon the specific source of data available to them (i.e., interview, psychometric, and observation, respectively) following a typical diagnostic study. The three categories of "ratable," "questionable," and "unratable" were designated for this purpose. To facilitate an analysis of judgments, the categories were assigned differential weights and the item values were summed across judges within each discipline. The ratable category was given a weight of 2, the questionable category a weight of 1, and the unratable category 0 weight.

The items were then analyzed for inter- and intradiscipline ratability with the dual purpose of sampling items judged ratable by every clinician and items presumably judged ratable on the basis of the specific source of data available to each discipline. Intradiscipline ratability was based on an item's receiving unit weight by each clinician and a weighted value greater than 50% of the possible total value for that discipline. For example, if each psychologist ($N = 4$) judged an item questionable, or, if three psychologists judged an item questionable and the fourth judged it unratable, the weighted value for that item would be either 4 or 3, respectively. The percentage value, however, would in neither case exceed 50% of the possible value of 8. For any item to be accepted, *ergo*, there must have been at least questionable agreement among judges and the item must have received a weighted value of 5 or more.

An analysis of judgments revealed that 213 items had received a maximum value of 26, indicating complete staff agreement as to their ratability. Ten items were judged either questionable or unratable by *every* clinician and received a weighted value of 13 or less. An example of an unratable item was: "Has a wish to be passively manipulated genitally."

Furthermore, 65 items showed weighted values indicating intra- but not interdiscipline ratability, and were presumably judged ratable on the basis of data peculiar to each discipline. Of this item total, 30 were ratable by caseworkers, 20 by psychologists, and 15 by group workers. The following are examples of these items for each discipline.

Casework: "Gets a good deal of satisfaction out of being able to control what goes on in the home."

Psychology: "Has a wish to kill people who thwart self in any way."

Groupwork: "Thrives on social (peer) group competition."

The remaining 331 items varied in value from 13 to 25. Because such a large number of these items were apparently of acceptable ratability (i.e., values between 20 and 25) it was arbitrarily decided to adhere to a stringent criterion for selection in order to reduce the size of the pool. Thus, only those items were retained for further screening which received a maximum ratability value of 26, and appeared ratable on the basis of data peculiar to each discipline. To provide equal representation for the latter, 15 items were sampled from each discipline. This procedure reduced the pool to 258 items. These items were sorted in the final screening process described below.

Before concluding this section it is important to note that no step was taken to insure the ratability of the items by the MMPI judges. This proce-

dure was purposely omitted in order to develop a unified pool appropriate for child guidance populations and usable by child guidance personnel. When MMPI judges were subsequently asked to judge items for ratability from the MMPI alone, approximately 10% of the items were uniformly judged unratable. When consideration is given to the 15 unratable items peculiar to each discipline, any bias is canceled out. Thus it can probably be concluded that the items were not prejudiced in favor of any data or discipline.

Screening Items for Interpatient Variability

The final step in the item selection procedure was designed to increase the discriminating power of the item pool. This phase involved an attempt to filter the pool of "Barnum algae" by eliminating all items that could be made to fit mothers and children largely or wholly by virtue of their triviality (Paterson, 1951). This issue has been discussed in greater detail by Meehl (1954), and is considered by others (Cronbach, 1953; Duker, 1958; Goodling & Guthrie, 1956; Halbower, 1955) to be an essential step in item selection. *Ergo,* to incorporate such discrimination insurance into the design, nine clinic therapists were requested to compile *Q*-sort descriptions of mothers and children then undergoing treatment. All therapists were experienced staff clinicians.

To insure a heterogeneous patient sample, five child therapists and four caseworkers were asked to submit a list of patients seen in treatment for a minimum of 10 hours. From this list the investigator chose at random four mothers, one child whose primary complaint was academic deficiency, and four children from each of the following diagnostic categories: psychotic disorder; psychoneurotic disorder; personality disorder; and psychophysiologic, autonomic, and visceral disorder. The 258 items were then sorted by therapists into an 11-category, semiforced, rectangular distribution. The therapists were instructed to sort a minimum of 15 items into each category according to the degree to which the items were characteristic of the person being described. The nine *Q* sorts which resulted from this procedure were used for the express purpose of eliminating items which showed little variability in category placement.

An item analysis was done as follows: the category value for each item placement (i.e., 1, 2, 3 . . . 11) was tabulated for all nine subjects. A "variance score" representing the number of contiguous categories of descriptiveness was computed for each of the 258 items. The process of item elimination was carried out for mothers and children jointly. All items with variance values less than 6 (i.e., varied over less than half the distribution) for mothers, children, or both, were eliminated. The following are examples of discarded items.

Mother: "Utilizes rigid, constrictive, suppressive, emotional control." "Has inferiority feelings."

Child: "Passivity (latent or manifest)." "Tends to be self-defensive; anticipates being attacked and criticized."

Mother and Child: "Utilizes repression as a defense mechanism." "Is concerned with own adequacy as a person, either at a conscious or unconscious level."

Final Q Array

The *Q* array in its final form consisted of 135 items which were believed to have the following attributes: reasonable representation of the broad domain of personality, appropriate for adults and children of both sexes, composed of pertinent clinical information, ratable by each member of the clinic staff, and adequate interpatient variability, i.e., relatively free of "Barnum-effect."

Twenty-five percent of the pool contained nonpathological statements. Of these, 9 items were neutral and 25 items were positive with regard to normal adjustment. Approximately 25% of the pool was judged to be comprised of genotypic item content. The 135 items retained in the final pool are presented in Appendix B.

Mother and Child Stereotype Descriptions

There is mounting evidence to the effect that, when provided minimal data regarding class membership (e.g., sex, age, education, etc., or that a person has patient status) judges can make personality predictions that exceed chance expectancy (Caldwell, 1958; Duker, 1958; Gage, 1953; Halbower, 1955; Hathaway, 1956; Kostlan, 1954; Sines, 1957). This issue has been discussed by Gage and Cronbach (1955) who conceptualize (predictive) accuracy scores as divided into two components: "stereotype accuracy" and "differential accuracy."

> Where the former refers to a judge's ability to predict the pooled responses of a given category of persons, the latter refers to ability to differentiate among persons within the category. . . . When accuracy is scored directly, no distinction can be made between the two components which contribute to the judge's success (p. 417).

To obviate such a state of affairs, the writers advocate obtaining at least two scores: (*a*) ability to predict the typical behavior in the next-larger class to which the person belongs (e.g., an "average" patient), and (*b*) ability to predict how a person deviates from the norm of this class. Testimony that accuracy scores warrant such an analysis is provided by a series of recent PhD dissertations at the University of Minnesota (Duker, 1958; Halbower, 1955; Sines, 1957). In each of these studies, the predictive efficiency of

judges' patient descriptions was compared with the accuracy of their patient stereotypes, and in many instances the stereotype was found to be the more accurate predictor.

To control for stereotype accuracy in the present study, each clinician who derived diagnostic *Q* sorts was asked to describe (by *Q* sort) his conception of the "average child," and the "average mother" referred to a child guidance clinic for evaluation and treatment. To obtain an estimate of stereotype reliability, each clinic psychologist compiled a resort approximately 9 months after his first stereotype description. Every clinician completed his stereotype descriptions before assignment of study subjects. Stereotype sorts, as well as all subject sorts, were forced into a quasi-normal, 11-category distribution. The latter was implemented to facilitate the numerous computations required.

The stereotype mother and child descriptions provided optimal chance estimates against which "differential accuracy" scores (i.e., subject predictions) could be compared. Two such comparisons were made for each subject in the study. First, each clinician's mother or child description was compared with his own stereotype for each mother and/or child described by him. Second, each clinician's mother or child description was compared with one of four composite clinic stereotypes (i.e., "mean average subject"). The latter were derived by fractionally omitting one stereotype and forming a composite of the remaining three. It was thus possible to compare the relative accuracy of MMPI-based descriptions and clinic-based descriptions, with a clinic-based composite stereotype which was free of each clinician's base rate bias.

Results

Reliability of Clinical Evaluators

Therapists' Child Criterion Descriptions. Intrasort reliability estimates (coefficients of stability) were computed from the resort descriptions of five clinic therapists who had compiled four or more criterion descriptions during the study. The number of days between sort and resort ranged from 2 to 14, and averaged 10. The coefficients were .93, .86, .85, .85, and .82. The five coefficients were transformed into z' scores, averaged, and converted to a mean coefficient of .87. All coefficients were significantly greater than zero beyond the .001 level.

Psychologists' Child Diagnostic Descriptions. Reliability estimates for child diagnostic descriptions were obtained from each clinic psychologist. Two clinical psychologist trainees repeated sorts of their final case 2 and 3 days following their first *Q*-sort descriptions. These coefficients were .80 and .79 for the 2 and 3 day sorts, respectively. Sort-resort correlations were also computed for each of five children assigned to staff psychologists for treatment. In each case, the psychologist who served as diagnostician (and who compiled diagnostic descriptions), also served as therapist and compiled the criterion description for the same child. For each child, the second description was separated from the first by approximately 3 months. The correlations obtained were .73, .62, .62, .56, and .55. The mean coefficient calculated through z transformation was .62. Again, all correlations were reliably greater than zero at well beyond the .001 level. Whereas the latter comparisons were made primarily to reflect the degree to which psychologists changed their impressions of children as a function of therapeutic contact with them, the coefficients can also be interpreted as estimates of the absolute minimum stability that might be expected of the *Q* array over an extended period of time.

Child Stereotype Descriptions. Four clinic psychologists and seven nonclinic MMPI judges derived child stereotype descriptions. Although only a fraction of the 55 possible intercorrelations were computed, it appeared that they would range from .38 to .62, and average about .52. The latter estimates were calculated by intercorrelating the four stereotype descriptions derived by clinic psychologists and then correlating each of the seven judges' stereotypes with the one clinic stereotype that most agreed with the other three. Accordingly, the child stereotype derived by Psychologist C was correlated with the seven judges' stereotypes. The coefficients obtained were .62, .56, .52, .51, .49, .46, and .38. Table 18 reports the intercorrelations for the four clinic psychologists.

In addition to the reliability estimates reported above, intrasort correlations were computed for child stereotype descriptions

TABLE 18

INTERSORTER RELIABILITY OF PSYCHOLOGISTS' CHILD STEREOTYPE DESCRIPTIONS

Psychologist	A	B	C[a]
B	.53		
C[a]	.53	.54	
D[a]	.44	.39	.52

[a] Third-year clinical psychologist trainee.

TABLE 19

INTERSORTER RELIABILITY OF CASEWORKERS' MOTHER STEREOTYPE DESCRIPTIONS

Caseworker	W	X	Y
X	.43		
Y	.61	.62	
Z	.60	.43	.49

which were resorted by the clinic psychologists. The time interval between sort and resort was approximately 9 months. The correlations obtained were .83, .70, .65, and .60. The mean coefficient computed through z transformation was .70. All of the above correlations were reliably greater than zero well beyond the .001 level.

Caseworkers' Mother Descriptions. Because of the excessive demand made upon caseworkers' time, no specific step was taken to secure repeat sorts of their diagnostic or criterion descriptions. Although estimates of (maximum) reliability were not available for either of these descriptions, a measure of minimum stability can be inferred from the correlations between them. The first and second mother descriptions were in most cases separated in time by approximately 3 months. The 42 sort-resort correlations ranged from .30 to .84. The mean coefficient computed through z transformation was .65. All of these correlations were significantly greater than zero beyond the .001 level.

It is of interest to note that the average coefficient obtained by correlating the four caseworkers' diagnostic and criterion mother descriptions (.65) approximates that similarly obtained by correlating the two psychologists' diagnostic and criterion child descriptions (.62). In both instances, the sort-resorts were separated in time by approximately 3 months.

Mother Stereotype Descriptions. Reliability estimates of mother stereotypes were calculated in a manner similar to that of the child stereotypes reported above. Four clinic caseworkers and five nonclinic MMPI judges derived "average mother" descriptions. Although only a fraction of the 36 possible intercorrelations were computed, it seemed that they would range from .25 to .62 and average about .52. The latter estimates (*i.q.* the child stereotypes) were computed by intercorrelating the four stereotypes compiled by the clinic caseworkers and then correlating each of the five MMPI judges' stereotypes with the one clinic stereotype that most agreed with the other three. Accordingly, the mother stereotype compiled by Caseworker Y was correlated with each of the five judges' descriptions. The coefficients obtained were .58, .58, .52, .46, and .25. Table 19 reports the intersorter correlations of four clinic caseworkers. All coefficients were reliably greater than zero beyond the .001 level.

A comparison of the correlations obtained from the child stereotype descriptions with those obtained from the mother stereotypes,

TABLE 20

INTRASORTER RELIABILITY OF JUDGES' MMPI-BASED MOTHER AND CHILD DESCRIPTIONS

Judge[a]	Days between Sorts	Mother	Child
I[b]	14	.75	.75
II[b]	21	.81	.75
IV	14	.85	.79
V[c]	7	.75	.88
VI[c]	14	.63	.45
All sorts		.77	.75

[a] Excluding Judges III and VII who derived *Q*-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] MMPI expert.
[c] MMPI neophyte.

showed that the median coefficients of both groups were the same (.52). Moreover, with exception of the judge's stereotype which correlated only .25 with the modal description, the range of the two distributions was similar (child .38 to .62; mother .43 to .62).

Judges' MMPI-Based Descriptions. Reliability estimates of mother and child MMPI-based descriptions were obtained from each judge. Five judges compiled resorts of mothers from MMPI profiles of mothers and of children from MMPI profiles of parents in a manner identical to that of their original mother and child descriptions. The intrasort reliability estimates for these descriptions are reported in Table 20.

It can be seen from Table 20 that the correlations for mothers ranged from .63 to .85 and averaged .77. For children, the correlations ranged from .45 to .88 and averaged .76. All coefficients were reliably greater than zero well beyond the .001 level.

Results of Testing the Hypotheses

Hypothesis I

Findings related to the first two hypotheses are presented in Tables 21 through 26. Tables 21 and 24 present means and standard deviations of MMPI scales for the two clinic samples and for two Minnesota general population normative groups selected for comparison. Since both groups differed reliably in the variance of several scales, the difference between means was computed by the approximation method given by Cochran and Cox (1957).

Hypothesis I posited personality variable differences between mothers of clinic patients and general population female adults. The results of testing this hypothesis appear in Tables 21, 22, and 23. Table 21 reports *T* score means and raw score means and standard deviations of MMPI clinical and validity scales for the clinic sample of 48 mothers and for 315 female adults of the general Minnesota population. Examination of Table 21 shows that *L* and *Pd* scales

TABLE 21

MMPI Validity and Clinical Scale Scores of Mothers of Child Guidance Clinic Patients and of General Population Female Adults

Scale	Mothers (N = 48)	Mothers (N = 48)		Females (N = 315)			
	T Scores	Raw Scores		Raw Scores			
	Mean	Mean	*SD*	Mean	*SD*	*t*	*F*
L	50.5	4.4	1.99	4.3	2.63	.43	1.74**
F	50.9	3.8	2.81	3.5	3.13	.64	1.24
K	56.7	15.5	4.58	12.1	5.07	4.64***	1.23
Hs[a]	54.5	15.3	5.13	13.1	4.88	2.74**	1.10
D	58.2	23.5	5.04	19.3	5.18	5.33**	1.05
Hy	60.7	24.9	6.13	18.8	5.66	6.34**	1.17
Pd[a]	61.1	23.3	5.32	18.4	4.40	5.95***	1.46*
Mf	49.7	36.3	4.82	36.5	4.83	.21	1.00
Pa	55.7	9.8	2.89	7.9	3.32	4.15***	1.33
Pt[a]	54.8	28.1	5.47	25.2	6.06	3.34***	1.23
Sc[a]	54.3	25.5	5.64	22.6	6.50	3.12**	1.33
Ma[a]	47.7	15.2	4.82	16.1	4.11	1.24	1.38
Si	54.9	29.7	8.75	25.0	9.58	3.35**	1.19

[a] *K*-corrected scales.
* Significant between .10 and .05 levels.
** Significant at .05 level.
*** Significant at .01 level.

yield F ratios significant at the .10 level or less when a two-tailed test is applied. Moreover, clinic mothers showed reliably higher mean scores on 9 of the 13 scales tested. All clinical scales with the exception of *Mf* and *Ma* were found reliably higher than the norm at the .05 level or beyond. The largest differences occurred on *Hy* and *Pd*. And, as might be expected, the *Hy* and *Pd* differences were also prominent in the frequency of clinical scale high points, the frequency of T scores equal to or greater than 70, and the mean MMPI code of clinic mothers' profiles.

Table 22 reports the percentage frequency of clinical scale high points as observed among the total number of females in both groups. The MMPI normative data used for this comparison were taken from Hathaway and Meehl (Welsh & Dahlstrom, 1956, p. 141). Since *Mf* and *Si* scale data were not reported by these authors, the mothers' high point codes were adjusted (the second high point was used) to facilitate a comparison with the norm. Table 22 shows reliable differences between the clinic and normative groups on two of eight clinical scales. Clinic mothers show a significantly higher percentage frequency of *Hy* and *Pd* high points, and a reliably lower frequency of no high point than is found in the general population. Moreover, *Hy* and *Pd* appeared as high points in approximately 50% of mothers' profiles as compared with a frequency of 16% in profiles of female adults. The high point which appeared most frequent among mothers was *Hy* (27%), whereas the most frequent high point in the normative group is *Ma* (13%).

Table 23 reports percentages of mothers and normals whose T scores equal or exceed 70 on each of eight clinical scales. Again, normative frequencies were not available for *Mf* and *Si*. Table 23 shows T scores equal to or greater than 70 (abnormal scores) occur with about equal frequency (1 to 3%) in the general population, and that approximately 14% of females in the normative group obtain a score of 70 or greater on at least one clinical scale. Unfortunately, the population percentages were too small to permit a statistical test of individual scale differences. The total percentage difference was computed, however, and indicates that a reliably greater number of mothers obtain abnormal scores than might be expected of a representative sample of the general pop-

TABLE 22

FREQUENCY OF MMPI CLINICAL SCALE HIGH POINTS OF MOTHERS OF CHILD GUIDANCE CLINIC PATIENTS AND OF GENERAL POPULATION FEMALE ADULTS

Scale	Mothers (N = 48) %	Females (N = 360) %	*CR*
Hs	4.2	8.6	–[a]
D	16.7	12.5	.81
Hy	27.1	7.5	4.30**
Pd	20.8	8.3	2.74*
Pa	12.5	7.2	–
Pt	2.1	6.4	–
Sc	2.1	6.1	–
Ma	8.3	13.0	–
No high point	6.2	25.8	3.01*

[a] Critical ratios were not computed for these scales as qN was less than 5 (McNemar, 1949, p. 77).
* Significant at .006 level.
** Significant at .002 level.

TABLE 23

FREQUENCY WITH WHICH MMPI CLINICAL SCALE T SCORES EQUAL OR EXCEED 70 AMONG PROFILES OF MOTHERS OF CHILD GUIDANCE CLINIC PATIENTS AND GENERAL POPULATION FEMALE ADULTS

Scale	Mothers (N = 48) %	Females (N = 360) %	*CR*
Hs	0	2.8	–[a]
D	4.2	2.8	–
Hy	14.6	1.1	–
Pd	14.6	.8	–
Pa	4.2	1.4	–
Pt	2.1	1.9	–
Sc	0	1.4	–
Ma	2.1	1.7	–
Total %[b]	41.8	13.9	4.82*
Total %[c]	50.0	13.9	6.10*

[a] Critical ratios were not computed for these scales as qN was less than 5 (McNemar, 1949, p. 77).
[b] Excluding *Mf* and *Si* scales.
[c] Including *Mf* and *Si* scales (in the clinic sample only).
* Significant at .002 level.

ulation. The percentages of abnormal profile scores were 50 and 14 for mothers and normals, respectively. Interestingly enough, the two scales which show the highest frequency of abnormal scores in the sample (*Pd* and *Hy*) have the lowest frequency as abnormal scores in the normative group.

It is also noteworthy that *Hy* and *Pd* were found to occupy the first and second positions among mothers' and fathers' profiles of a sample of parents referred to the Washburn Memorial Clinic in Minneapolis (Hanvik & Byrum, 1959). The mean rank profile codes[7] of both samples were 34260-71859 and 3462780159 for the study clinic and Washburn samples, respectively. A rank order correlation between these codes yielded a coefficient of .93 indicating extreme similarity of mothers in the two clinic populations.

An examination of clinical scale high points revealed five code patterns, which combined represented 70% of all profiles in the clinic mother sample.

Codes 13 and 31: This pattern, commonly referred to as the "conversion V" or "hysteroid valley," was found among 29% of the mothers' profiles (cf. 33% of mothers' profiles in the Washburn sample). Some follow-up data collected after 6 to 9 months of treatment[8] indicated that, regardless of elevation, a majority of *cases* in which the mothers' code was of this pattern responded favorably to treatment.

Codes 24 and 42, and 34 and 43: As compared with the 34 and 43 pattern, mothers in the 24-42 group were given slightly better ratings by therapists following 6 months of treatment. Neither group, however, presented a very favorable prognosis. Mothers with the 34-43 pattern canceled many interviews and seldom remained in treatment beyond 3 months. Any elevation on Scale O (*Si*), especially if equal to or greater than a *T* score of 60, was a malignant sign. Hanvik and Byrum found a high *Pd* minus *Ma* differential frequently obtained by mothers who were situationally maladjusted and reacting adversely and with extreme hostility to a portion of the social environment, and who presented problems of severe marital conflict. The average *Pd* minus *Ma* score of their mother group was 9.7. This differential was also computed for mothers of the present study and found to average 12.8. An examination of the *Pd* minus *Ma* differential for the groups rated for response to treatment revealed scores of 2.8, 15.6, and 20.4 for the improved, partially improved, and unimproved groups, respectively. The relative magnitude of the score for the partially improved group is of especial interest when the criteria for the rating are known. This group was comprised of cases in which the child's therapist had judged the child improved, whereas the caseworker had judged the mother unable to form a relationship. Thus, as might be expected, the *Pd* minus *Ma* differential of the partially improved group more closely approximates that of the unimproved than of the improved group.

Codes 20 and 02: The most frequent pattern associated with the partially improved group was 20-02. Scale 0, especially if associated with Scale 2 in second position,

[7] The coding system employed is that described by Hathaway (1947). The clinical scales are identified by the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 0 in order from left to right in the usual profile arrangement (1 *Hs*, 2 *D*, 3 *Hy*, etc.).

[8] At the time criterion descriptions were compiled each therapist independently rated each child and mother in terms of whether they were able to form a "therapeutic relationship." At the close of the study each caseworker rated each case in terms of response to treatment. The latter ratings were made from 6 to 9 months after therapy had begun. From an analysis of these ratings, three groups were identified as improved, partially improved, and unimproved. The improved group was comprised of 15 cases in which both mother and child were able to form a relationship and subsequently were able to respond favorably to treatment. The partially improved group was comprised of 18 cases in which the child was able to form a relationship and to respond to treatment, but the mother was not. The unimproved group included 15 cases in which neither the mother nor child was able to relate to a therapist nor respond to treatment, or the case terminated treatment against clinic advice. In several instances somewhat arbitrary decisions had to be made regarding assignment to groups; however, each decision was made independent of knowledge of the MMPI.

occurred among mothers' profiles of half the cases judged partially improved, whereas only 6% of mothers who obtained this pattern received a different rating.

Code 9: Although relatively infrequent as a high point and seldom elevated in any profile, Code 9 was nevertheless the best single index of improvement. Every case in which the mother's profile coded 9 high was judged improved.

Hypothesis II

Hypothesis II posited personality variable differences between fathers of clinic patients and general population male adults. Tables 24, 25, and 26 report data testing this hypothesis. In Table 24, *T* score means and raw score means and standard deviations of MMPI clinical and validity scales are presented for 32 clinic fathers and for 226 adult males of the general Minnesota population. Table 24 shows *F, D, Pa,* and *Sc* yielded *F* ratios significant at the .05 level or less. Moreover, for 8 of the 13 scales tested, scores of fathers were significantly higher than the mean of the normative group at the .05 level or beyond. The only clinical scales which failed to yield reliable mean differences were *Pa, Ma,* and *Si.* The largest differences occurred for *D, Hy, Pd,* and *Pt* which were all reliable well beyond the .01 level. No scale mean was observed significantly lower than the mean of the normative group.

In Table 25 the frequency of clinical scale high points as observed among males in both groups is presented in percentage form. Since *Mf* and *Si* data were not available for adult males, the clinic codes were redistributed and the second high point used. An examination of Table 25 shows that clinic fathers, like clinic mothers, showed a significantly higher frequency of *Hy* high point and reliably lower frequency of no high point (zero frequency) than is found among adults in the normative group. The most frequent high point among fathers was *D,* which was significantly greater than the norm at the .002 level. Whereas *Hy* and *Pd* appeared as high

TABLE 24

MMPI Validity and Clinical Scale Scores of Fathers of Child Guidance Clinic Patients and of General Population Male Adults

Scale	Fathers (N = 32)	Fathers (N = 32)		Males (N = 226)			
	T Scores	Raw Scores		Raw Scores			
	Mean	Mean	*SD*	Mean	*SD*	*t*	*F*
L	50.0	4.0	2.04	4.1	2.89	.05	2.01
F	51.2	4.4	1.84	3.9	4.24	.39	5.32**
K	55.0	16.3	4.66	13.4	5.66	3.11*	1.46
Hs[a]	56.7	14.0	5.04	11.3	3.90	2.84*	1.67
D	62.0	21.7	6.35	16.6	4.18	4.37**	2.31**
Hy	59.5	21.7	5.50	16.5	5.51	5.04**	1.00
Pd[a]	59.5	23.2	4.12	19.3	4.11	5.05**	1.00
Mf	54.6	22.8	4.83	20.4	5.13	2.58*	1.13
Pa	50.0	8.4	2.49	8.1	3.56	.69	2.03*
Pt[a]	56.8	26.3	5.29	22.9	4.88	3.37**	1.17
Sc[a]	54.8	24.8	7.41	22.3	5.21	2.07*	2.02*
Ma[a]	51.4	17.6	4.11	17.0	3.87	.71	1.18
Si	51.8	26.8	9.68	25.0	9.58	.94	1.02

[a] *K*-corrected scales.
* Significant at .05 level.
** Significant at .01 level.

TABLE 25

FREQUENCY OF MMPI CLINICAL SCALE HIGH POINTS OF FATHERS OF CHILD GUIDANCE CLINIC PATIENTS AND OF GENERAL POPULATION MALE ADULTS

Scale	Fathers ($N = 32$) %	Males ($N = 258$) %	*CR*
Hs	15.6	10.4	.89
D	25.0	6.6	3.50**
Hy	21.9	7.0	2.69*
Pd	15.6	11.6	.66
Pa	6.2	7.7	–[a]
Pt	0	6.7	–
Sc	6.2	4.8	–
Ma	9.4	17.4	1.15
No high point	0	23.7	3.11**

[a] Critical ratios were not computed for these scales as qN was less than 5 (McNemar, 1949, p. 77).
* Significant at .006 level.
** Significant at .002 level.

points in approximately 50% of mother profiles, *Hy* and *D* appeared as high points in 47% of father profiles. It is notable that *Pt*, which appears elevated with *D* for a large number of psychiatric males, was nonexistent as a high point among clinic fathers.

Table 26 reports the percentage frequency of *T* scores equal to or greater than 70 among the clinic and normative male groups. An examination of Table 26 shows that abnormal scores occur with frequencies up to about 5% among general population males, and that 15% of normals obtain *T* scores of 70 or greater on at least one clinical scale. Although the frequencies of abnormal scores on individual scales were too small for statistical comparisons, the total percentage difference between groups was computed and found significant at the .002 level. Thus, 44% of clinic fathers and 50% of clinic mothers obtained abnormal scores on one or more scales as compared with 15% of males and 14% of females of the normative groups. It is notable that the most frequent high point of fathers (*D*) is among the most infrequent high points of normals, whereas the most frequent high point of male adults (*Ma*) has zero frequency among the clinic group.

A comparison of the above data for fathers with findings reported by Hanvik and Byrum revealed little similarity. The mean rank profile codes were 2435718906 and 5342691870 for the study clinic and Washburn samples, respectively. A rank order correlation between codes yielded a coefficient of .59, which fell short of significance.

An examination of clinical scale high points yielded four code patterns, which combined represented 85% of all profiles in the clinic father sample.

Codes 20 and 02: Thirty-seven percent of clinic fathers obtained profiles in which Scales 2 and 0 were found in prominent positions. This pattern was also found frequent among profiles of clinic mothers (see discussion *supra*) and in general was also characteristic of the average profile of the total father sample. Profiles with 20-02 *T* values of 70 and above occurred with a frequency of about 70%. The 20-02 pattern, more often than not, indicated an unfavorable prognosis. However, since elevation itself was negatively related to improvement, the latter to some extent might be expected. It should be noted, moreover,

TABLE 26

FREQUENCY WITH WHICH MMPI CLINICAL SCALE *T* SCORES EQUAL OR EXCEED 70 AMONG PROFILES OF FATHERS OF CHILD GUIDANCE CLINIC PATIENTS AND GENERAL POPULATION MALE ADULTS

Scale	Fathers ($N = 32$) %	Males ($N = 258$) %	*CR*
Hs	9.4	4.3	–[a]
D	12.5	1.9	–
Hy	9.4	.4	–
Pd	6.2	1.9	–
Pa	0	.4	–
Pt	0	1.5	–
Sc	3.1	.4	–
Ma	0	4.3	–
Total %[b]	40.6	15.1	4.75*
Total %[c]	43.7	15.1	4.80*

[a] Critical ratios were not computed for these scales as qN was less than 5 (McNemar, 1949, p. 77).
[b] Excluding *Mf* and *Si* scales.
[c] Including *Mf* and *Si* scales (in the clinic sample only).
* Significant at .002 level.

that both 2 and 0 (*D* and *Si*) have been found sensitive to elevations with age (Brozek, 1955; Calden & Hokanson, 1959; Sopchak, 1958). And, since clinic fathers were significantly older than general population males, some elevation would be expected as a function of age alone. Contrary to expectation, however, an elevation of Scale 2 did *not* indicate "counseling readiness" for either parent in the clinic sample.

Codes 13 and 31: This pattern was found to occur in 19% of clinic fathers' profiles (cf. 24% of fathers' profiles of the Washburn sample). The 13-31 pattern of fathers, unlike that of mothers, was associated with case unimprovement.

Code 4: In 16% of fathers' profiles *Pd* was the most prominent scale. The Code 4 pattern was characterized by secondary elevations on *K* and Scale 9 with Scales 5 and 0 generally down. Fifty percent of profiles with this pattern showed one or more scales with *T* values of 70 or above. In contrast to the finding for mothers, *Pd* elevation for fathers was most often associated with case improvement. However, rarely did fathers with this pattern participate in clinic treatment. Despite the relatively high incidence of *Pd* among mothers *and* fathers, in only one case did *Pd* appear as the high point for both. It would be of interest to know the incidence of *Pd* as a high point among parents of delinquents!

Code 95: Scales 9 and 5 appeared in first and second position, respectively, in 16% of clinic fathers' profiles. In no instance, however, was the 95 pattern associated with abnormal scores on any scale, nor was *Pd* prominent in any profile. Among these profiles, Scales 1 and 2 were invariably down, and often the latter appeared below a *T* value of 50. More often than not, the 95 code for fathers was associated with case improvement (*q.v.* Code 9 for mothers *supra*).

Hypothesis III

Results which pertain to Hypotheses III, IV, and V were analyzed with nonparametric statistical techniques. Although parametric statistics are commonly used in *Q* research, the assumptions on which they are based can seldom if ever be legitimately made. For example, the product-moment correlational technique necessitates the assumption that the variables (*Q*-sort statements) have been randomly sampled from a universe of statements descriptive of the personality domain. However, since the parameters of personality are largely unknown, and since most statements are selected with some definite expediency in mind (e.g., ratability), a departure from the statistical model is unavoidable. The most common circumvention of the problem has been to interpret the coefficients "as if" the sampling assumption could be made. A more defensible alternative, however, would be to interpret the coefficient as an *index* of similarity only (i.e., as means of ordering data) and to process this index with statistics of the nonparametric model. In the present study both the Sign and Binomial tests were used for this purpose.

Hypothesis III posited that judges could derive personality descriptions of mothers from blind readings of their MMPI profiles which would correlate more highly with caseworkers' therapy-based (criterion) descriptions than with descriptions of the same mothers derived by the same caseworkers after only 3 interview hours. Findings which pertain to this hypothesis are reported in Table 27.

Table 27 presents the mean correlations with caseworkers' diagnostic and criterion descriptions of judges' mother predictions. These coefficients were obtained by correlating each judge's MMPI-based description with similar descriptions compiled by caseworkers following brief and more extended contact. The average coefficients were computed through *z* transformation. The procedure for testing the hypothesis was as follows: The correlation of Judges' MMPI-based descriptions with caseworkers' diagnostic and criteria descriptions were compared for each judge; if the judge's description correlated more highly with the criterion than with the diagnostic description, a plus was tallied; if the judge's description correlated more highly with the diagnostic than with the criterion descrip-

TABLE 27

MEAN CORRELATIONS OF JUDGES' MMPI-BASED MOTHER DESCRIPTIONS WITH CASEWORKERS' DIAGNOSTIC DESCRIPTIONS AND CASEWORKERS' CRITERION DESCRIPTIONS

Judges[a]	*Q* Sorts	Diagnostic	Criterion	Criterion *r* higher	
	N	*r*	*r*	*N*	*p*
I[b]	8	.20*	.32***	7	.03
II[b]	9	.21*	.41***	8	.02
IV	8	.21*	.31***	6	.14
V	9	.28***	.38***	8	.02
VI	8	.24**	.34***	7	.03
All sorts	42	.23**	.36***	36	.001

[a] Excluding Judges III and VII who derived *Q*-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] MMPI expert.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

tion, a minus was tallied; if the judge's description correlated the same with both the diagnostic and the criterion description (*i.e.*, there was no difference in the magnitude of the coefficient), a tie was recorded; separate tallies were made for each judge and then combined for different comparisons.

An examination of Table 27 shows that the total mean correlation of judges' MMPI-based descriptions with caseworkers' diagnostic descriptions was .23, and that the mean correlation of the former with caseworkers' criteria descriptions was .36. The two coefficients are reliably greater than zero at the .01 and .001 levels, respectively. In Column 5 of Table 27 the number of plus signs (i.e., criterion *r* higher than diagnostic *r*) is entered for each judge. Column 6 reports one-tailed probabilities associated with the occurrence of values in Column 5. It can be seen from Column 6 that for four out of five judges the individual probabilities lead to a rejection of the null hypothesis of no difference in the comparative accuracy of the two descriptions at the .03 level or beyond. Furthermore, in 36 out of 42 total comparisons, judges' descriptions agreed more closely with the criterion. For such an extreme split, the Binomial test (corrected for continuity) yields a very low associated probability under the null hypothesis of no difference between the two descriptions. Since the hypothesis predicted the direction of the difference, the region of rejection is one-tailed and the probability value is .001.

Hypothesis IIIa. Hypotheses IIIa and IIIb posited that judges' MMPI-based descriptions of mothers would exceed chance expectancy. Since the hypotheses are related, the findings are presented jointly and appear in Table 28. In Column 3 of Table 28 are entered the mean correlations with the criteria of mother stereotype descriptions compiled by five MMPI judges. The average correlation of these descriptions was .21 which is reliably greater than zero at the .05 level. For comparative purposes, the mean correlations with the criteria of judges' MMPI-based descriptions are presented in Column 9. The average correlation of these descriptions was .36 which is reliably greater than zero at the .001 level. In Column 4 are entered the number of plus signs (i.e., MMPI-based *r* higher than stereotype *r*) for each judge. Column 5 reports the one-tailed probabilities associated with the occurrence of values in Column 4. From these data it can be seen that for three of the five judges the individual probabilities lead to a rejection of the null hypothesis of no difference between the stereotype and MMPI-based descriptions at the .09 level or beyond. Although the Sign test did not yield probabilities greater than chance for two of the judges, the average correlation of their MMPI-based descriptions was greater in magnitude than that of their stereotypes in both instances. The important test of Hypothesis IIIa, however, involved the combined performance of five judges. Table 28 shows that in 31 out of 42 such comparisons, the accuracy of judges' MMPI-based descriptions surpassed that of their own steretoypes. Since the hypothesis was directional, the region of rejection is one-tailed and the probability value is .002. On the basis of combined data, Hypothesis IIIa found support.

TABLE 28

MEAN CORRELATIONS WITH THE CRITERION OF JUDGES' MMPI-BASED MOTHER DESCRIPTIONS, JUDGES' STEREOTYPE MOTHER DESCRIPTIONS, AND CASEWORKERS' COMPOSITE STEREOTYPE MOTHER DESCRIPTIONS

Judge[a]	Q Sorts	Judges' Stereo.	Mother r higher		Composite Stereo.	Mother r higher		Judges' MMPI
	N	r	N	p	r	N	p	r
I[b]	8	.09	7	.03	.35***	4	.64	.32***
II[b]	9	.29***	6	.25	.33***	7	.09	.41***
IV	8	.14	7	.03	.18*	5	.36	.31***
V	9	.24**	7	.09	.23**	7	.09	.38***
VI	8	.27**	4	.64	.30***	5	.36	.34***
Totals	42	.21*	31	.002	.28***	28	.02	.36***

[a] Excluding Judges III and VII who derived Q-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] MMPI expert.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

Hypothesis IIIb. Columns 6, 7, and 8 of Table 28 report data which pertain to Hypothesis IIIb. In Column 6 of Table 28 are entered the mean correlations with the criteria of composite mother steretotype descriptions compiled by three of four clinic caseworkers. The average correlation of these descriptions was .28 which is reliably greater than zero at the .001 level. In Column 7 are presented the number of plus signs (i.e., MMPI-based *r* higher than composite stereotype *r*) for each judge. Column 8 contains the one-tailed probabilities associated with the occurrence of values in Column 7. An examination of these data for individual judges shows that in at least three instances the probability values do *not* lead to rejection of the null hypothesis of no difference between the accuracy of the descriptions. Again, however, the important test involved the total performance of five judges. It can be seen from Table 28 that in 28 out of 42 such comparisons, the accuracy of judges' MMPI-based descriptions surpassed that of caseworkers' composite stereotypes. The one-tailed probability associated with the occurrence of this split led to a rejection of the null hypothesis of no difference at the .02 level. On the basis of the combined performance of five judges, Hypothesis IIIb found support.

Hypothesis IIIc. Hypotheses IIIc and IIId posited that caseworkers' diagnostic descriptions of mothers would exceed chance expectancy. Since both hypotheses predicted greater agreement between caseworkers' diagnostic and criterion descriptions than between their stereotype and criterion descriptions, the findings for each are considered jointly. In Column 3 of Table 29 are entered the mean correlations with the criteria of mother stereotype descriptions derived by four clinic caseworkers. These coefficients ranged from .27 to .49, and averaged .41. All are reliably greater than zero at the .001 level. For comparative purposes, the mean correlations with the criteria of caseworkers' diagnostic descriptions (i.e., their average sort-resort correlations) are presented in Column 9. These coefficients ranged from .62 to .69 and averaged .65. An inspection of the correlations in Columns 3 and 9 shows considerable variability among the former and relative stability among the latter. The discrepancy in the magnitude of caseworkers' own stereotype "loadings" was somewhat disconcerting. Coefficients as high as .49

(Caseworkers W and Z) indicate, *ceteris paribus*, that a sizable percentage of variance associated with the reliability of two caseworkers' descriptions of mothers can also be accounted for by reference to their prevailing conception of the "average clinic mother."

In Column 4 of Table 29 the number of plus signs (i.e., diagnostic r higher than stereotype r) is reported for each caseworker. Column 5 presents one-tailed probabilities associated with values in Column 4. These data show that for each of the four caseworkers the associated probabilities lead to a rejection of the null hypothesis of no difference in the accuracy of the stereotype and diagnostic descriptions at the .06 level or beyond. Moreover, in 37 out of 42 total comparisons the accuracy of diagnostic descriptions surpassed that of the stereotypes. The probability value associated with such an extreme split is .001. These findings support Hypothesis IIIc.

Hypothesis IIId. In Columns 6, 7, and 8 of Table 29 are reported data which pertain to Hypothesis IIId. The coefficients in Column 6 were obtained by correlating the combined stereotype (composite) description of three caseworkers with each criterion description of the fourth and then averaging the correlations (through z transformation) to obtain a mean coefficient. The four average composite stereotypes thus obtained ranged from .18 to .34 with a mean of .28. Three of the four coefficients were reliably greater than zero at the .001 level. One was reliably greater than zero at the .05 level. Although the mean composite stereotype coefficients in Column 6 are as variable as the mean individual stereotype coefficients in Column 3, the magnitude of the latter exceeds that of the former in every instance. This discrepancy in the accuracy of composite and individual stereotype "loadings" suggests that individual caseworkers ascribe a greater percentage of "average mother" variance to criterion mother descriptions than is warranted.

In Columns 7 and 8 of Table 29 the number of plus signs (i.e., diagnostic r higher than composite stereotype r) and the probabilities associated with these values are entered for each caseworker. The data for individual caseworkers lead to a rejection of the null hypothesis of no difference in the comparative accuracy of the composite stereotype and diagnostic descriptions at the .02 level or beyond. Moreover, the probability associated with 41 positive of 42 total values is .001. Thus, on the basis of *both* individual and combined results Hypothesis IIId was supported.

Hypothesis IV

Hypothesis IV posited that judges' descriptions of *children* derived from *parent* personality test data would better predict

TABLE 29

MEAN CORRELATIONS WITH THE CRITERION OF CASEWORKERS' DIAGNOSTIC, STEREOTYPE, AND COMPOSITE STEREOTYPE MOTHER DESCRIPTIONS

Caseworker	*Q* Sorts	Caseworkers' Stereo.	Mother r higher		Composite Stereo.	Mother r higher		Caseworkers' Resort
	N	r	*N*	*p*	r	*N*	*p*	r
W	15	.49**	11	.06	.33**	14	.001	.63**
X	11	.37**	11	.001	.28**	11	.001	.70**
Y	10	.27**	9	.01	.18*	10	.001	.62**
Z	6	.49**	6	.02	.34**	6	.02	.65**
Totals	42	.41**	37	.001	.28**	41	.001	.65**

* Significant at .05 level.
** Significant at .001 level.

criterion descriptions compiled by clinic therapists than diagnostic descriptions compiled by clinic psychologists.

Table 30 presents data which pertain to Hypothesis IV. In Column 4 of Table 30 are entered mean correlations with the criteria of descriptions of children extrapolated from parent MMPI profile data. Entries in Column 3 were obtained by correlating and averaging each judge's descriptions with similar descriptions derived by clinic psychologists.

An inspection of Column 3 shows that the mean correlations of judges with psychologists ranged from .09 to .32 and averaged .22. Two of the five mean coefficients are reliably greater than zero at the .001 level; two are reliable at the .05 level, and one is not significant. In Column 4 are entered the mean correlations (validity coefficients) of judges with therapists. These mean coefficients ranged from .06 to .41 and averaged .28. Three are reliably greater than zero at the .001 level, one at the .01 level, and one is not significant. An inspection of the mean correlations for each judge reveals that, with exception of Judge I, all were in closer agreement with therapists than with psychologists.

TABLE 30

MEAN CORRELATIONS OF JUDGES' EXTRAPOLATED CHILD DESCRIPTIONS WITH PSYCHOLOGISTS' DIAGNOSTIC DESCRIPTIONS AND THERAPISTS' CRITERION DESCRIPTIONS

Judge[a]	*Q* Sorts	Diagnostic	Criterion	Criterion *r* higher	
	N	*r*	*r*	*N*	*p*
I[b]	7	.09	.06	3	.50
II[b]	8	.20*	.24**	5	.36
IV	8	.27***	.36***	7	.03
V	8	.32***	.41***	7	.03
VI	7	.19*	.29***	4	.50
All sorts	38[c]	.22*	.28***	26	.02

[a] Excluding Judges III and VII who derived *Q*-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] MMPI expert.
[c] Excluding four cases due to inaccurate criterion.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

In Column 5 of Table 30 the number of plus signs (i.e., criterion *r* higher than diagnostic *r*) are entered for each judge. Column 6 presents the one-tailed probabilities associated with the occurrence of values in Column 5. It can be seen from Column 6 that for two out of five judges the individual probability values lead to a rejection of the null hypothesis of no difference in the comparative accuracy of the two descriptions at the .03 level. For three judges the individual probability values fail to reach an acceptable level. The important test of the hypothesis, however, concerned the total performance of five judges. An inspection of this total in Row 6 shows that in 26 out of 38 comparisons, judges' descriptions agreed more closely with the criterion. For such a split the Binomial test yields a relatively low probability under the null hypothesis of no difference. Since the hypothesis predicted the direction of difference, the region of rejection is one-tailed and the probability value is .02. Hypothesis IV was supported.

Hypothesis IVa. Hypotheses IVa and IVb posited that judges' descriptions of children extrapolated from parent personality test data would exceed chance expectancy. Since the two hypotheses are related, the findings are presented jointly and appear in Table 31. In Column 3 of Table 31 are entered the mean correlations with the criteria of child stereotype descriptions compiled by five MMPI judges. These correlations ranged from .15 to .24 and averaged .20. All but one mean coefficient (.15) is reliably greater than zero at the .05 level or beyond. For comparative purposes the mean correlations with the criteria (validity coefficients) of judges' extrapolated child descriptions are entered in Column 9. These coefficients ranged from .06 to .41 and averaged .28. Three of the five coefficients are significantly greater than zero at the .001 level, one is reliable at the .01 level, and one is not significant. An examination of these coefficients shows that, again with exception of Judge I, all mean correlations with the criteria of judges' child descriptions were greater in magnitude than were correspond-

ing coefficients based upon judges' own stereotypes.

In Column 4 of Table 31 are entered the number of plus signs (i.e., MMPI-based r higher than stereotype r) for each judge. Column 5 reports the one-tailed probability values associated with the occurrence of signs in Column 4. It can be seen that for three out of five judges the individual probability values lead to a rejection of the null hypothesis of no difference in the accuracy of the stereotype and extrapolated child descriptions at the .06 level or beyond. As the important test of Hypothesis IVa concerned the total performance of five judges, these data are reported separately in Row 6. Row 6 shows that in 29 out of 38 total comparisons, the accuracy of judges' extrapolated child descriptions surpassed that of their own stereotypes. For such a split, the Binomial test yields a probability value of .001. Hypothesis IVa found support.

Hypothesis IVb. In Columns 6, 7, and 8 of Table 31 data are entered which pertain to Hypothesis IVb. In Column 6 of Table 31 are given the mean correlations with the criteria of composite child stereotype descriptions derived by three of four clinic psychologists. These mean coefficients ranged from .16 to .28 and averaged .23. Two coefficients are reliably greater than zero at the .001 level, one at the .01 level, one at the .05 level, and one is not significant. Again, with exception of Judge I, all mean correlations with the criteria of composite child stereotypes were lower in magnitude than corresponding validity coefficients of judges' MMPI-based descriptions.

In Column 7 of Table 31 are entered the number of plus signs (i.e., MMPI-based r higher than composite stereotype r) for each judge. In Column 8 are reported one-tailed probability values associated with signs in Column 7. An examination of these values for individual judges shows that in only one instance (Judge V) does the associated probability lead to a rejection of the null hypothesis of no difference in the accuracy of the two descriptions. Row 6 presents data concerning the combined performance of five judges. It can be seen from these data that in 26 out of 38 comparisons the accuracy of judges' child descriptions surpassed that of psychologists' composite stereotypes. The probability associated with this split is .02. Thus, on the basis of the combined performance of five judges, Hypothesis IVb found support.

TABLE 31

MEAN CORRELATIONS WITH THE CRITERION OF JUDGES' EXTRAPOLATED CHILD DESCRIPTIONS, JUDGES' STEREOTYPE CHILD DESCRIPTIONS, AND PSYCHOLOGISTS' COMPOSITE STEREOTYPE CHILD DESCRIPTIONS

Judge[a]	Q Sorts	Judges' Stereo.	Child r higher		Composite Stereo.	Child r higher		Judges' MMPI
	N	r	N	p	r	N	p	r
I[b]	7	.24**	2	.77	.24**	2	.77	.06
II[b]	8	.15	6	.14	.16	6	.14	.24**
IV	8	.18*	7	.03	.18*	6	.14	.36***
V	8	.21*	8	.004	.27***	7	.03	.41***
VI	7	.24**	6	.06	.28***	5	.23	.29***
Totals	38[c]	.20*	29	.001	.23**	26	.02	.28***

[a] Excluding Judges III and VII who derived Q-sort descriptions of adolescents only. The results of these descriptions can be found elsewhere (Marks, 1959b).
[b] MMPI expert.
[c] Excluding four cases due to inaccurate criterion.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

Hypothesis V

Hypothesis V posited that psychologists' diagnostic evaluation-based descriptions would better predict the criterion than descriptions of the same children derived from personality test data of their parents. Findings which relate to this hypothesis are presented in Table 32. In Table 32 are given the mean correlations with the criteria of descriptions of children derived by clinic psychologists and by MMPI judges. Entries in Column 4 were obtained by correlating and averaging each psychologist's diagnostic descriptions with criteria descriptions derived by clinic therapists. These coefficients ranged from .38 to .42 and averaged .39. All are reliably greater than zero at the .001 level. In Column 3 are entered mean correlations with the criteria of child descriptions extrapolated by judges from MMPI profiles of parents. These coefficients ranged from .21 to .35 and averaged .29. Two coefficients are reliably greater than zero at the .001 level and two at the .05 level. Both mean coefficients are reliably greater than zero at the .001 level. An examination of the coefficients in Column 4 shows that the descriptions compiled by psychologist trainees agreed more closely with the criteria than did those of staff psychologists. Moreover, the coefficients of trainees' descriptions show greater discrepancy when compared with judges than do coefficients of descriptions derived by clinic staff. In Column 5 are given the number of plus signs (psychologists' r higher than judges' r) for each psychologist. Column 6 presents the probabilities associated with corresponding values in Column 5. Unfortunately, the small size of the sample for trainees precluded application of the Sign test for their data. The probability values associated with the performance of staff psychologists failed to reach an acceptable level. An inspection of Row 5 shows that, *even* when combined, the data fail to yield a probability value that would lend support to the hypothesis. Thus, on the basis of *both* individual and combined results, Hypothesis V was rejected.

TABLE 32

MEAN CORRELATIONS WITH THE CRITERION OF PSYCHOLOGISTS' CHILD DIAGNOSTIC DESCRIPTIONS AND JUDGES' EXTRAPOLATED CHILD DESCRIPTIONS

Psychologist	*Q* Sorts	Judges' Child	Psychologists' Child	Psychologists' r higher	
	N	r	r	*N*	*p*
A	17	.30**	.38**	10	.31
B	8	.35**	.39**	5	.36
C[a]	4	.22*	.42**	3	[b]
D[a]	4	.21*	.40**	2	[b]
Totals	33[c]	.29**	.39**	20	.15

[a] Third-year clinical psychologist trainee.
[b] Small size of *N* precluded application of the Sign test.
[c] Excluding four cases due to inaccurate criterion and five cases in which psychologist served as therapist.
* Significant at .05 level.
** Significant at .001 level.

Hypothesis Va. Hypotheses Va and Vb posited that psychologists' descriptions of children would surpass chance expectancy. Data pertaining to these hypotheses appear in Table 33. In Column 3 of Table 33 are entered the mean correlations with the criteria of child stereotype descriptions derived by clinic psychologists. These coefficients ranged from .15 to .27 and averaged .24 (cf. .20 for judges). One coefficient is reliably greater than zero at the .001 level, two at the .01 level, and one is not significant. For comparative purposes the mean correlations with the criteria of psychologists' diagnostic descriptions are entered in Column 9. These coefficients ranged from .38 to .42 and averaged .39 (cf. .28 for judges).

In Column 4 of Table 33 the number of plus signs (i.e., diagnostic r higher than stereotype r) are reported for each psychologist. Column 5 gives the one-tailed probability values associated with signs in Column 4. Again, the small size of the sample for trainees precluded application of the Sign test for their data. The probability values associated with the performance of staff psychologists revealed that only the descriptions of Psychologist A were reliably greater than chance. The combined total of the four psychologists appears in Row 5. It can be seen from Row 5 that in 27 out of

TABLE 33

MEAN CORRELATIONS WITH THE CRITERION OF PSYCHOLOGISTS' DIAGNOSTIC, STEREOTYPE, AND COMPOSITE STEREOTYPE CHILD DESCRIPTIONS

Psychologist	*Q* Sorts	Psychologists' Stereo.	Child *r* higher		Composite Stereo.	Child *r* higher		Psychologists' Child
	N	*r*	*N*	*p*	*r*	*N*	*p*	*r*
A	17	.26**	13	.02	.22*	12	.07	.38***
B	8	.22**	6	.34	.27***	7	.03	.39***
C[a]	4	.27***	4	[b]	.18*	4	[b]	.42***
D[a]	4	.15	4	[b]	.27***	3	[b]	.40***
Totals	33[c]	.24**	27	.001	.23**	26	.001	.39***

[a] Third-year clinical psychologist trainee.
[b] Small size of *N* precluded application of the Sign test.
[c] Excluding four cases due to inaccurate criterion and five cases in which psychologists served as therapist.
* Significant at .05 level.
** Significant at .01 level.
*** Significant at .001 level.

33 comparisons the diagnostic descriptions surpassed the stereotypes. The Binomial yielded a probability value of .001. On the basis of the combined data for four psychologists, Hypothesis Va found support.

Hypothesis Vb. In Columns 6, 7, and 8 of Table 33 are entered data which relate to Hypothesis Vb. The coefficients in Column 6 were obtained by correlating the combined (composite) stereotype descriptions of three psychologists with the criteria and then averaging the correlations to obtain mean coefficients. Correlations of four average composite stereotypes with the criteria ranged from .18 to .27 and averaged .23 (cf. .23 for judges). Two of the four coefficients were reliably greater than zero at the .001 level and two at the .05 level. In Columns 7 and 8 the number of plus signs (i.e., diagnostic *r* higher than composite stereotype *r*) and the probabilities associated with these values are entered for each psychologist. Again, the small size of the sample for trainees precluded an application of the Sign test for their data. The probability values associated with the performance of staff psychologists leads to a rejection of the null hypothesis of no difference in the comparative accuracy of the composite stereotype and diagnostic descriptions at the .007 level or beyond. The probability associated with the combined plus values of all psychologists is .001. Hypothesis Vb was supported.

DISCUSSION[9]

To recapitulate, the study was designed, in part, to assess the efficiency of various kinds of data available to child guidance practitioners for consideration in making decisions. The sources of data were both test and nontest. The procedure employed permitted comparisons of inferences drawn from the MMPI with inferences drawn from clinical interviews of mothers and from psychological examinations of children.

Accuracy of Clinicians' Patient Descriptions

Considerable attention was given to the selection of items used in the study. The essential features of the design (i.e., the sampling of both "normal" and pathological statements from the broad domain of personality, and obtaining therapist [criterion] ratings of item "pertinence") insured a heterogeneous pool of presumable content-valid items. Thus, having "built" reliability

[9] A more extensive discussion of the findings can be found in the original paper.

into the pool, it was not surprising to find that subsequent estimates of stability, especially of the child criterion, were of substantial size. The most noteworthy finding among the various subset resorts obtained, was the relative (and absolute!) magnitude of the mean stability estimates of child descriptions by psychologists and judges. These coefficients, and the number of days separating sort from resort were .77 for psychologists and .75 for judges, for 2 and 14 days, respectively. Of lesser interest, perhaps, but of import to an issue discussed below were the stability estimates of mother descriptions. These descriptions, separated in time by 3 months, and by 14 days, yielded mean coefficients of .65 and .77 for caseworkers and judges, respectively.

Turning now to validity, certain findings, among which were estimates of the accuracy of the criterion, question the credibility of psychotherapists in matters of criterion import. Placing overly high confidence in the "truth" value of therapists' judgments on the basis of reputed experience, expertness, skill, or the like, e.g., "enlisting the cooperation of . . . training analysts" (Silverman, 1959, p. 19), while a fashionable expedient, may prove to be as much a source of professional security as the confidence (once) placed in diagnostic techniques! It is puzzling, indeed, why some writers (e.g., Kogan, 1954, p. 666) are so confident in their appraisal of the criterion problem.

Another source of disillusionment may be found in viewing reliability as a surrogate for validity. For example, the coefficient of stability (intrasort reliability) can be manipulated to yield an estimate of the upper limit of validity. Such an estimate, however, turns out to be empirically unrealistic, since *no* validity coefficients even approximate this magnitude. It seems obvious that the criterion is sorely in need of validation itself.

Unfortunately, the situation is not much improved by eliminating the major source of error in the criterion. The findings are all too clear on this point. For, despite a substantially high mean stability coefficient of .87, and, despite a correction for inaccuracy of the criterion, therapist-psychologist agreement over the total sample of 33 children averaged only .39, and ranged from .05 to .67. Thus, even though the design *precluded* uncontaminated therapist-psychologist descriptions, less than 20% of the predicted variance could be accounted for by the latter's descriptions. Moreover, only 22% of the variance could be accounted for when the error variance was partialed out, i.e., when a complete correction for attenuation was made. These findings offer little cause for rejoicing. What they suggest, is that diagnosticians (psychologists) and therapists, in dealing with the same patients, in the same clinic, under presumably optimal conditions, and sharing the same information, cannot communicate with accuracy the salient features of patients' personalities. In short, they operate from incongruent frames of reference, which, has the net effect of rendering the psychological examination of children, if not the diagnostic process itself, a somewhat indefensible procedure.

Furthermore, the findings offer little support for the commonly held belief that the collection, combination, and integration of various sources of data (i.e., a procedure actually carried out in the clinical situation) enables the clinician to make statements about personalities which are significantly more accurate than those made on the basis of blind interpretation of much limited data. Consequently, there is little support for the notion that knowledge about, and "understanding of," are linearly related, or that "understanding of," and "treatment success," are necessarily related! Attention should be called to the fact that at least one therapist had very little knowledge about (as defined) and presumably, therefore, "understanding of," any of his patients. Nevertheless, all of his patients were judged to have improved (by standards other than the therapist's, of course).

The obvious possibility comes to mind that, given certain *kinds* of optimally weighted data, any surplus contributes more error than true variance. This issue has

been discussed in detail by Meehl (1957) and needs no elaboration here. Suffice it to say that, even though the optimal weights are yet largely unknown, the evidence favoring the use of certain kinds of information (and disfavoring the use of other kinds) is becoming quite clear. For example, notwithstanding the relative superiority (note: relative, not absolute) of the MMPI as compared with its competitors, the clinician might take more seriously the findings of Hathaway (1956) and others which suggest that such classificatory items as age, sex, socioeconomic status, education, occupational status, etc. (i.e., "minimal data") have predictive power capable of clinically important generalizations (e.g., Caldwell, 1958; Kostlan, 1954; Sines, 1957; Taft, 1955). Evidence to this effect can be found in the present study by comparing the relative magnitude of the various data-based Q correlations with that of the stereotypes.

In summary, when judged in absolute terms, the accuracy of the various data-based predictions ranges from substantial to poor. Most impressive is the stability of the child criterion (suggesting that the Q array can be sorted reliably). Somewhat less impressive is the magnitude of caseworkers' sort-resort correlations. The average coefficient over the total of 42 cases was only .65, suggesting that whatever information serves as the basis for decisions, is of only moderate reliability. The MMPI-based mother descriptions yielded an average validity coefficient of .36, which is moderate. Perhaps of greater interest is the fact that in 36 out of 42 comparisons ($p = .001$), judges' MMPI-based descriptions of mothers correlated more highly with caseworkers' descriptions after 10 hours than after only 3. The convergence results of MMPI-based descriptions of children, though in the direction of those of mothers, yielded relatively low mean coefficients. Most discouraging, were the low congruency findings for psychologists and therapists who, for the most part, shared similar data (*pro dolor*).

Accuracy of Clinicians' Stereotype Descriptions

A second consideration in evaluating efficiency, requires a comparison of the accuracy of predictions made from the data relative to that of prediction made on the basis of chance. The findings pertaining to such comparisons in the present study offer several points of interest.

As expected, the average correlations of each of the various subset stereotype descriptions with the criterion were all significantly greater than zero. However, as not expected, the magnitude of the average correlations with the criterion of caseworkers' and of psychologists' individual stereotypes ($r = .41$ caseworkers, $r = .24$ psychologists) surpassed that of independent composite descriptions ($r = .28$ caseworkers, $r = .22$ psychologists). It will be recalled that the composite stereotype descriptions were derived from the four child stereotypes compiled by psychologists and the four mother stereotypes compiled by caseworkers and, after fractionally omitting one from each set of four, by forming a composite of the remaining three. Accordingly, four composite child stereotypes and four composite mother stereotypes (consisting of three descriptions each) were obtained for the express purpose of comparing each data-based prediction with a presumably more accurate stereotype (a more stringent estimate of chance) that was free of each clinician's stereotype bias. Both theoretically and empirically, when compared with an individual stereotype, a composite stereotype (an actuarial description) should yield a more accurate prediction. What then is the implication when it does not? In the present study, especially for caseworkers, the obvious implication is that a sizable percentage of the variance associated with the accuracy of their descriptions of mothers can also be accounted for by reference to their prevailing conception of the "average clinic mother." In other words, failing to discriminate, they describe mothers "pretty much the same." That this is unwarranted (i.e., that mothers are not "pretty much the same"), can be in-

ferred from the magnitude of the discrepancy between the individual and composite mother stereotype coefficients (viz., between an $r = .28$ and an $r = .41$).

In summary, clinicians can predict better than chance not only on the basis of their own perceptions of the average clinic child and mother, but on the basis of a presumably more accurate perception comprised of a consensus of psychologist and caseworker opinion.

Major Implications

An important implication of the findings is that the clinical utility of any diagnostic device, insofar as it is a function of the clinician-as-an-instrument variable, demands that the device be validated both within the setting in which it is used and by the clinician who chooses to use it. The clinician simply cannot legitimately assume that the validation of data by their "clinical usefulness" in his own particular setting or that *even* the formal validation of data in another similar setting, is sufficient in the absence of objective evidence to the effect that the clinician's own use of the data "pays off." It cannot be seriously contended that the present findings support the continued use of certain projective techniques which on the basis of their use with children are of such low or moderate validity; and hence, make little appreciable contribution to patient welfare (cf. their use with adults) (Kostlan, 1954; Little & Shneidman, 1959; Silverman, 1959; Sines, 1957).

The state of affairs which quite possibly exists in many child guidance clinics today requires an evaluation not only of the devices employed in the psychological evaluation of children, but of the diagnostic process itself. Since the latter is considered the *sine qua non* for planning effective treatment and since the present findings seriously question the efficiency of the process, it is disheartening to contemplate the effect upon patients that results from decisions based upon such an endeavor, in addition to the sheer waste of clinical time. Insofar as clinicians (psychiatrists, social workers, and psychologists) are willing to question their premises and conclusions, psychologists should be willing to investigate the validity of these premises and conclusions. In short, the urgency of his therapeutic function should not deter the psychologist from his more important function—research.

Another important implication of the findings concerns the extension of psychological testing in general, and of MMPI testing in particular, to *parents* of child guidance patients. Despite the fact that psychological tests have been administered to children since the inception of the child guidance movement, it was only recently that the diagnostic testing of parents was even discussed programatically (Rosenzweig & Cass, 1954). One source of difficulty seems to stem from the alleged threat, both to the parent and to the caseworker, that parent testing seems to imply. Suffice it to say that the present study offered no evidence that *parents* felt threatened by such an endeavor. What it did suggest, however, was first that it might well be the clinician rather than the parent who feels threatened; and second, was that many parents require treatment which extends beyond the "basic structure of the casework process" (Perlman, 1953, p. 308) and, admittedly, casework techniques in child guidance are not entirely appropriate for the diagnosis *or* treatment of psychiatric disorders in adults (Lippman, 1956). Consequently, clinics should either offer such service or should refer such cases to more appropriate sources—assessment is inevitable.

Summary and Conclusions

The present study was designed to investigate personality characteristics of parents of child guidance patients and to assess the relative efficiency of various sources of data available to clinicians for consideration in making pretreatment decisions.

Forty-eight cases, representing 90% of consecutive referrals accepted for treatment between March 15, 1958 and July 31, 1958, comprised a normative group. Of these, 42 cases in which the mother and child remained in treatment for a minimum of 5

hours and in which the child was free of gross organic (cerebral) impairment and/or mental deficiency, comprised a validation group.

The average child was a Caucasian, 10-year-old boy in the fourth grade. His intelligence was average but he was referred to the clinic by the school because of academic deficiency. He was described by his parents as nervous and fearful, and was reported to have sleep disturbances and poor peer relationships. The average child was a Protestant and came from a home that had not been disrupted by death, separation, or divorce. His mother and father were 37 and 41 years of age, respectively. Both parents had received a tenth grade education. The child's father was employed in a skilled clerical, or business (Class III) occupation.

Case assignments, diagnostic methods, and treatment procedures followed established clinic routine, with the exception that MMPI testing was extended to include parents. Upon completion of diagnostic evaluations, psychologists and caseworkers compiled *Q*-sort personality descriptions of children and of mothers, respectively. The modal psychological examination of children consisted of an intelligence estimate, three personality measures, and some index of academic achievement. The modal evaluation of mothers consisted of three clinical interviews.

Testing the various hypotheses involved comparisons of MMPI data from clinic mothers and fathers with similar data from their respective normative groups, and comparisons of the relative accuracy of (*a*) *Q*-sort personality descriptions of children, derived by psychologists from psychological examination data of children; (*b*) *Q*-sort personality descriptions of children, derived by nonclinic judges from blind readings of the MMPI profiles of children's parents; (*c*) *Q*-sort child stereotype descriptions, derived by clinic psychologists from their own conception of the "average" clinic child; (*d*) *Q*-sort child stereotype descriptions, derived by nonclinic judges from their own conception of the "average" clinic child; (*e*) *Q*-sort child composite stereotype descriptions, derived by clinic psychologists from their combined conception of the "average" clinic child; (*f*) *Q*-sort personality descriptions of mothers, derived by clinic caseworkers from clinical interview data of mothers; (*g*) *Q*-sort personality descriptions of mothers, derived by nonclinic judges from blind readings of the MMPI profiles of mothers; (*h*) *Q*-sort mother stereotype descriptions, derived by clinic caseworkers from their own conception of the "average" clinic mother; (*i*) *Q*-sort mother stereotype descriptions, derived by nonclinic judges from their own conception of the "average" clinic mother; and (*j*) *Q*-sort mother composite stereotype descriptions, derived by clinic caseworkers from their combined conception of the "average" clinic mother.

The criteria with which the various descriptions were compared for predictive accuracy were *Q* sorts compiled by clinic therapists with whom the children and their mothers had had an average of 10 hours of interview contact. The design precluded uncontaminated *clinic* *Q*-sort descriptions. Criterion descriptions of mothers were compiled by the same caseworkers who had interviewed and had *Q* sorted the mothers diagnostically. Moreover, no attempt was made to deprive therapists of psychometric data about children, for such data are normally available to them and therapists' judgments are presumably founded to some extent upon information available in psychological reports. Only the MMPI was withheld from clinic personnel. The latter, in turn, was made available to seven nonclinic student and expert judges.

Two separate prediction tasks were required of the judges by the design. *Q*-sort descriptions of mothers were derived from blind readings of their MMPI profiles, and *Q*-sort descriptions of children were derived from MMPI profiles of their parents. An 11-category, forced, quasinormal distribution was used. Stereotype mother and child descriptions in addition to reliability data were obtained from every clinician.

The *Q* array was empirically derived by the clinic staff from approximately 2,000 items sampled from various sources. In its final form, the array consisted of 135 items

which had been judged a reasonable representation of the broad domain of personality, appropriate for children and adults of both sexes, composed of pertinent clinical information, ratable by each member of the clinic staff, and of adequate intersubject variability for both mother and child. Approximately 25% of the array contained genotypic items (75% phenotypic). Twenty-five percent of the array contained nonpathological statements. Of these, 7% were neutral and 18% were descriptive of optimal adjustment.

The findings support the following conclusions:

1. Parents of child guidance clinic patients differ from general population adults with respect to most personality variables measured by the MMPI.

When compared with general population females, clinic mothers show significantly higher scores on *K, Hs, D, Hy, Pd, Pa, Pt, Sc,* and *Si* with the strongest trends on *Hy* and *Pd*. Moreover, abnormal scores occur in 50% of clinic mothers' profiles as compared with percentage frequencies of 14 and 75 among normal and psychiatric groups, respectively. An examination of clinical scale high points reveals five code patterns (13-31, 24-42, 20-02, 34-43, and 9) which together represent 70% of all profiles in the clinic mother sample.

The findings for clinic fathers reveal scores significantly higher than general population males on *K, Hs, D, Hy, Pd, Mf, Pt,* and *Sc* with the strongest trends on *D* and *Hy*. Abnormal scores occur with a percentage frequency of 44 among fathers as compared with 15 and 75 among normal and psychiatric groups, respectively. An examination of clinical scale high points reveals four code patterns (13-31, 20-02, 4, and 59-95) which together represent 85% of all profiles in the clinic father sample.

In 77% of clinic families, at least one parent demonstrates symptoms and personality adjustment patterns which are commonly characteristic of psychiatric abnormality. Ratings of case movement reveal Codes 13-31 and 9 for mothers, and Codes 4 and 59-95 for fathers associated with rated case improvement, and Codes 24-42 and 34-43 for mothers, and Code 13-31 for fathers associated with rated case unimprovement. Code 20-02, for either parent and regardless of elevation, is associated with an unfavorable prognosis. Similarly, a high *Pd* minus *Ma* differential is a malignant sign. The average *Pd* minus *Ma* scores of mothers of cases rated improved, partially improved, and unimproved are 2.8, 15.6, and 20.4, respectively.

2. The MMPI, whether used with mothers as a source of data for generating inferences about mothers or whether used with parents as a source of data for generating inferences about children, is more efficient than the diagnostic process itself.

3. Personality descriptions of mothers derived by judges from blind readings of MMPI profiles of mothers correlate more highly with similar descriptions derived by caseworkers following 10 interview hours than with descriptions of the same mothers derived by the same caseworkers following only 3 interview hours.

4. Personality descriptions of children derived by judges from blind readings of MMPI profiles of children's parents are no less accurate than similar descriptions derived by clinic psychologists following completion of routine psychological examinations of the same children.

5. Clinicians (judges, psychologists, and caseworkers) can compile personality descriptions of mothers and of children which correlate more highly with a criterion than do stereotype descriptions compiled by the same clinicians.

6. Clinicians (judges, psychologists, and caseworkers) can compile personality descriptions of mothers and of children which correlate more highly with a criterion than do composite stereotype descriptions compiled by other clinicians.

7. A moderately high percentage of variance associated with the reliability of caseworkers' descriptions of mothers can also be accounted for by reference to caseworkers' own prevailing conception of the "average clinic mother."

8. Psychotherapists, if used as a criterion, are in need of validation themselves.

9. The collection, combination, and integration of various sources of data, i.e., what is actually done in the clinical situation, does not enable the clinician to make inferences about personality which are significantly more accurate than inferences based upon blind interpretation of much limited data.

10. The overall agreement between psychologists and therapists, in dealing with the same patients, in the same clinic, and sharing the same information, is discouragingly low. In short, they operate from incongruent frames of reference, which, has the net effect of rendering the psychological examination if not the diagnostic process itself an extravagant waste of clinical time.

11. In addition to engendering personality descriptions which are useful in treatment planning and in case disposition, MMPI code patterns yield practical actuarial data which predict response to clinic treatment.

12. The validity of any diagnostic device, insofar as it is a function of the clinician-as-an-instrument variable, demands that the device be validated both within the setting in which it is used and by the clinician who chooses to use it. The clinician cannot legitimately assume that the validation of data by their clinical usefulness in his own particular setting or that the formal validation of data in another similar setting is sufficient in the absence of evidence to the effect that his *own use* of the data "pays off."

REFERENCES

ACKERMAN, N. W. *Psychodynamics of family life.* New York: Basic Books, 1958.

ACKERMAN, N. W., & BEHRENS, MARJORIE L. A study of family diagnosis. *Amer. J. Orthopsychiat.*, 1956, **26**, 66–78.

BROZEK, J. Personality changes with age: An item analysis of the MMPI. *J. Geront.*, 1955, **10**, 194–206.

CALDEN, G., & HOKANSON, J. E. The influence of age on MMPI responses. *J. clin. Psychol.*, 1959, **15**, 194–195.

CALDWELL, A. B. Personality impressions and response prediction. Unpublished doctoral dissertation, University of Minnesota, 1958.

COCHRAN, W. G., & COX, GERTRUDE M. *Experimental design.* (2nd ed.) New York: Wiley, 1957.

CRONBACH, L. J. Correlations between persons as a research tool. In O. H. Mowrer (Ed.), *Psychotherapy theory and research.* New York: Ronald, 1953. Pp. 376–388.

DUKER, JAN. The utility of the MMPI atlas in the derivation of personality descriptions. Unpublished doctoral dissertation, University of Minnesota, 1958.

GAGE, N. L. Explorations in the understanding of others. *Educ. psychol. Measmt.*, 1953, **13**, 14–26.

GAGE, N. L., & CRONBACH, L. J. Conceptual and methodological problems in interpersonal perception. *Psychol. Rev.*, 1955, **62**, 411–422.

GOODENOUGH, FLORENCE, L., & ANDERSON, J. E. *Experimental child study.* New York: Century, 1931.

GOODLING, R., & GUTHRIE, G. Some practical considerations in *Q*-sort item selection. *J. counsel. Psychol.*, 1956, **3**, 70–72.

HALBOWER, C. C. A comparison of actuarial versus clinical prediction to classes discriminated by the Minnesota Multiphasic Personality Inventory. Unpublished doctoral dissertation, University of Minnesota, 1955.

HANVIK, L. J., & BYRUM, M. MMPI profiles of parents of child psychiatric patients. *J. clin. Psychol.*, 1959, **15**, 427–431.

HATHAWAY, S. R. A coding system for MMPI profiles. *J. consult. Psychol.*, 1947, **11**, 334–337.

HATHAWAY, S. R. Clinical intuition and inferential accuracy. *J. Pers.*, 1956, **24**, 223–250.

HATHAWAY, S. R. A study of human behavior: The clinical psychologist. *Amer. Psychologist*, 1958, **13**, 257–265.

HATHAWAY, S. R., & BRIGGS, P. F. Some normative data on new MMPI scales. *J. clin. Psychol.*, 1957, **13**, 364–368.

HOLLINGSHEAD, A. B., & REDLICH, F. C. Social *class and mental illness: A community study.* New York: Wiley, 1958.

HOLLIS, FLORENCE. Casework diagnosis—what and why? *Smith Coll. Stud. soc. Wk.*, 1954, **24**, 1–8.

HUNT, W. A., JONES, N. F., & HUNT, E. B. Reliability of clinical judgment as a function of clinical experience. *J. clin. Psychol.*, 1957, **13**, 377–378.

KOGAN, L. S. Research in clinical psychology. In L. A. Pennington & I. A. Berg (Eds.), *An introduction to clinical psychology.* New York: Ronald, 1954. Pp. 666.

KORMAN, M. An investigation of semantic differences among, and the inferential processes of, psychiatrists, psychologists and psychiatric social workers. Unpublished doctoral dissertation, University of Minnesota, 1957.

KOSTLAN, A. A method for the empirical study of psychodiagnosis. *J. consult. Psychol.*, 1954, **18**, 83–88.

KOSTLAN, A. A reply to Patterson. *J. consult. Psychol.*, 1955, **19**, 486.

LIPPMAN, H. S. *Treatment of the child in emotional conflict.* New York: McGraw-Hill, 1956.

LITTLE, K. B., & SHNEIDMAN, E. S. The validity of MMPI interpretations. *J. consult. Psychol.*, 1954, **18**, 425–428.

LITTLE, K. B., & SHNEIDMAN, E. S. Congruencies among interpretations of psychological test and anamnestic data. *Psychol. Monogr.*, 1959, **73**(6, Whole No. 476).

LIVSON, N. H., & NICHOLS, T. F. Discrimination and reliabiilty in *Q*-sort personality descriptions. *J. abnorm. soc. Psychol.*, 1956, **52**, 159–165.

MARKS, P. A. *Audio form of the MMPI.* Minneapolis: Author, 1959. (a)

MARKS, P. A. The validity of the diagnostic process in a child guidance setting: A multidisciplinary approach. Unpublished doctoral dissertation, University of Minnesota, 1959. (b)

MAAS, H. S., KAHN, A. J., & SUMNER, DOROTHY. Socio-cultural factors in psychiatric clinic services for children: A collaborative study in the New York and San Francisco metropolitan areas. *Smith Coll. Stud. soc. Wk.*, 1955, **25**, 1–90.

MCNEMAR, Q. *Psychological statistics.* New York: Wiley, 1949.

MEEHL, P. E. *Clinical versus statistical prediction.* Minneapolis: Univer. Minnesota Press, 1954.

MEEHL, P. E. Symposium on clinical and statistical prediction. *J. counsel. Psychol.*, 1956, **3**, 163–168. (a)

MEEHL, P. E. Wanted—a good cookbook. *Amer. Psychologist*, 1956, **11**, 263–272. (b)

MEEHL, P. E. When shall we use our heads instead of the formula? *J. counsel. Psychol.*, 1957, **4**, 268–273.

MEEHL, P. E. The cognitive activity of the clinician. *Amer. Psychologist,* 1960, **15,** 19–27.

MEEHL, P. E., & ROSEN, A. Antecedent probability and the efficiency of psychometric signs, patterns, or cutting scores. *Psychol. Bull.,* 1955, **52,** 194–215.

O'NEAL, PATRICIA, & ROBINS, L. N. The relation of childhood behavior problems to adult psychiatric status: A 30 year follow-up study of 150 subjects. *Amer. J. Psychiat.,* 1958, **114,** 961–969.

PATERSON, D. G. Character reading at sight of Mr. X according to the system of Mr. P. T. Barnum. Quoted in M. L. Blum & B. Balinsky (Eds.), *Counseling and psychology.* New York: Prentice-Hall, 1951. P. 47.

PATTERSON, C. H. Diagnostic accuracy or diagnostic stereotypy? *J. consult. Psychol.,* 1955, **19,** 483–485.

PATTERSON, G. R. A tentative approach to the classification of children's behavior problems. Unpublished doctoral dissertation, University of Minnesota, 1955.

PERLMAN, HELEN H. The basic structure of the case-work process. *Soc. serv. Rev.,* 1953, **27,** 308–315.

ROACH, J. L., GURRSLIN, O., & HUNT, R. G. Some social-psychological characteristics of a child guidance clinic caseload. *J. consult. Psychol.,* 1958, **22,** 183.

ROSENZWEIG, S., & CASS, LORETTA K. The extension of psychodiagnosis to parents in the child guidance setting. *Amer. J. Orthopsychiat.,* 1954, **24,** 715–722.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

SILVERMAN, L. H. A *Q*-sort study of the validity of evaluations made from projective techniques. *Psychol. Monogr.,* 1959, **73**(7, Whole No. 477).

SINES, L. K. An experimental investigation of the relative contribution to clinical diagnosis and personality description of various kinds of pertinent data. Unpublished doctoral dissertation, University of Minnesota, 1957.

SOPCHACK, A. L. Spearman correlations between MMPI scores of college students and their parents. *J. consult. Psychol.,* 1958, **22,** 207–209.

STEVENS, SYLVIA. An ecological study of child guidance intake. *Smith Coll. Stud. soc. Wk.,* 1954, **25,** 73–84.

SWENSEN, C. H. Empirical evaluations of human figure drawings. *Psychol. Bull.,* 1957, **54,** 431–466.

TAFT, R. The ability to judge people. *Psychol. Bull.,* 1955, **52,** 1–23.

WELSH, G. S., & DAHLSTROM, W. G. *Basic readings on the MMPI in psychology and medicine.* Minneapolis: Univer. Minnesota Press, 1956.

(Received June 16, 1960)

APPENDIX A

Behavior Rating Scale

(Total symptoms and referral complaints)

Behavior Rating	Total *N*	Total %	Behavior Rating	Total *N*	Total %
Unable to achieve in school	22	46	Excessive lying	4	8
Fearful	12	25	Lack of effort: listless, apathy	4	8
Nervous, anxious	12	25	Physical handicaps: hearing, vision, etc.	4	8
Sleep disturbances: nightmares, insomnia, etc.	12	25	Dreamy, day dreams	4	8
Poor peer relationships	12	25	Tics, habit spasms, mannerisms	4	8
Passive: overly conforming, unable to fight back	11	23	Stealing: delinquent	3	6
Tense, overly sensitive	11	23	Defiant, nonconforming, disobedient	3	6
Shy, withdrawn	10	21	Destructive	3	6
Enuretic, soiling	10	21	Sex difficulties	3	6
Cries easily	9	19	Temper tantrums, screaming	3	6
Psychosomatic symptoms	9	19	Excessive fantasy	3	6
Eating difficulties	9	19	Excessive fighting	2	4
School phobia	8	17	Truant	2	4
Perfectionistic	8	17	Fire setting	2	4
Bizarre behavior	7	15	Immature thumb sucking	2	4
Hyperactive, impulsive, unpredictable	7	15	Emotionally overreactive	2	4
Difficult to control, unmanageable	7	15	Suicide attempt	2	4
Unable to concentrate	6	13	Compulsive acts, obsessed with specific topics	2	4
Aggressive	5	10	Does not speak	2	4
Mischievous, provocative	5	10	Negativistic	2	4
Speech difficulties: articulation, phonetics, etc.	5	10	Runs away	2	4
			Demands attention	1	2

APPENDIX B

THE *Q* ARRAY

1. Reports difficulty in thinking (e.g., cannot concentrate).
2. Tends to be ruminative and overideational.
3. Obsessive thinking present.
4. Is perfectionistic: is compulsively meticulous.
5. Is socially extraverted (outgoing).
6. Manifests hypochondriacal tendencies, i.e., is excessively concerned about physical condition and functioning, is hypersentive to and overevaluates little pains and dysfunctions.
7. Is self-dramatizing: histrionic.
8. Is excitable.
9. Complains of weakness or is easy fatigued.
10. Has feelings of hopelessness.
11. Has a high aspiration level for self: is ambitious, wants to get ahead.
12. Judges self and others in conventional terms like "popularity," "the correct thing to do," "social pressures," etc.
13. Tends to arouse liking and acceptance in people.
14. Has a rapid personal tempo: thinks, talks, moves at a fast rate.
15. Is cheerful.
16. Is vulnerable to real or fancied threat: generally fearful, is a worrier.
17. Experiences difficulty in giving orders or making demands and requests of others.
18. Has "diagnostic" insight: awareness of the *descriptive* features of own behavior. (Examples: that certain symptoms are neurotic, that one is not liked by others, that one tends to distort in certain ways; that one is depressed, that one shows poor judgment in such-and-such ways, that one underachieves.)
19. Exhibits good heterosexual adjustment.
20. Is demanding: tends to take the attitude "the world owes me a living," "I have a right to be taken care of," etc.
21. Is unpredictable and changeable in behavior and attitudes.
22. Is egocentric, self-centered, selfish: seeks need-gratifications ("how will this affect me?") with little regard for the happiness and well-being of others.
23. Appears to be poised, self-assured, socially at ease.
24. Is defensive about admitting psychological conflicts: tries to avoid revealing self as having psychological conflicts and emotional distresses.
25. Delusional thinking is present.
26. Has grandiose ideas (extreme is delusions of grandeur).
27. Seeks out and tries to relate to parent figures.
28. Is resentful.
29. Exhibits psychotic tendencies.
30. Exhibits depression (manifest sad mood).
31. Tends to delay or avoid action: fears committing self to any definite course, is indecisive, vacillating.
32. Is evasive.
33. Is irritable.
34. Tends to transfer blame.
35. Demands sympathy from others.
36. Behaves considerately towards others.
37. Is argumentative.
38. Presents self as being physically, organically sick.
39. Is suggestable: overly responsive to other people's evaluations rather than own.
40. Tends to be rebellious and nonconforming.
41. Thinks and associates in unusual ways: has unconventional thought processes (extreme is illogical, confused, or bizarre).
42. Is apathetic.
43. Has inner conflict about sexuality (distinguish from reality problems in this area).
44. Has inner conflict about emotional dependency (distinguish from reality problems in this area).
45. Utilizes *acting out* as a defense mechanism.
46. Reacts to frustration intropunitively (i.e., punishes self).
47. Utilizes *regression* as a defense mechanism.
48. Exhibits evidence of narcissism (latent or manifest).
49. Has conflicts about giving.
50. Has good verbal-cognitive insight into own personality structure and dynamics, and has a real "feeling" for these insights. Insight not defended against by isolation-intellectualization.
51. Has a *need* to affiliate with others: i.e., to form friendships and associations, to greet and converse sociably with others, to join various groups, etc.
52. Has inner conflicts about self-assertion (distinguished from reality problems in this area).

53. Fears loss of control, feels need to keep rigid check on own emotional responses: cannot "let go" even when appropriate.

54. Utilizes *projection* as a defense mechanism.

55. Has a *need* to achieve: to overcome obstacles, to exercise power, to strive to do something difficult as well and as quickly as possible (this is an elementary ego need which may alone prompt action or be fused with any other need).

56. Is retentive: has a *need* to retain possession of things; to refuse to give or lend; to hoard; to be frugal, economical, and miserly.

57. Places value on intellectual and cognitive activities, skills, and attitudes.

58. Utilizes *rationalization* as a defense mechanism.

59. Defenses are fairly adequate in relieving psychological distress.

60. Tends toward overcontrol of needs and impulses: binds tensions excessively, delays gratification unnecessarily.

61. Utilizes *intellectualization* as a defense mechanism.

62. Life has included rewarding socialization experiences.

63. Would be organized and adaptive when under stress or trauma.

64. Has a resilient ego-defense system: has a safe margin of integration, adequate self-control.

65. Values wealth or material possessions and judges self and others in terms of them.

66. Gets appreciable "secondary gain" from symptoms (i.e., symptoms function to get person out of painful, difficult, or stressful situations in a socially acceptable way or are otherwise rewarding via their manipulations of external-relations).

67. Characteristically pushes and tries to stretch limits: sees what can be gotten away with.

68. Is self-defeating: places self in an obviously bad light.

69. Is nervous, tense in manner: trembles, sweats, or shows other *manifest* signs of anxiety.

70. Is distrustful of people in general: questions their motivations.

71. Is readily dominated by others: is submissive.

72. Spends a good deal of time in personal fantasy and daydreams: fictional speculations.

73. Keeps people at a distance: avoids close interpersonal relationships.

74. Is sensitive to anything that can be construed as a demand.

75. Accepts others as they are: is not judgmental.

76. Is able to sense other person's feelings: is an intuitive, empathetic person.

77. Is protective of those close to self (placement of this item expresses behavior ranging from overprotection through appropriate nurturance to laissez-faire, unstructuring attitudes).

78. Is critical, not easily impressed, skeptical.

79. Genotype has psychopathic features.

80. Genotype has schizoid features.

81. Genotype has hysteroid features.

82. Genotype has paranoid features.

83. Genotype has obsessive-compulsive features.

84. Shows concern over reputation.

85. Feels there is social stigma attached to clinic contact.

86. Is a serious person who tends to anticipate problems and difficulties, and to look at the "dark side" of things.

87. Is concerned about the qualifications of various staff members.

88. Seems unable to express own emotions in any modulated, adaptive way.

89. There are many "positives" in this case.

90. Presents a favorable prognosis.

91. Is overanxious about minor matters and reacts to them as if they were real emergencies.

92. Is a shy, anxious, and inhibited person.

93. Resorts to escape *into* fantasy.

94. Has unresolved Oedipal problems.

95. Tends to be flippant both in word and gesture.

96. Is open and frank in discussing problems.

97. Has a *need* to think of self as an unusually self-sufficient person.

98. Has a tenuous hold on reality.

99. Has shown ability to talk about conflicts in most areas.

100. Would be threatened by interpretations given early in therapy.

101. Is suffering from feelings of rejection.

102. Has a relatively mature superego.

103. Tends not to become involved in things: passively resistant.

104. Is stereotyped and unoriginal in approach to problems.

105. Undercontrols own impulses: acts with insufficient thinking and deliberation.

106. Gets along well in the world as it is: is socially appropriate in own behavior, keeps out of trouble (to be considered as conceptually separate from person's intrapsychic state).

107. Is tearful and/or cries openly.

108. Emphasizes oral pleasures: is self-indulgent.

109. Is consciously guilt-ridden: self-condemning, self-accusatory.

110. Doesn't seem to be particularly afraid of anything.

111. Is tense, high-strung, jumpy: has an overreadiness to respond with startle or apprehension to unexpected stimulation.

112. Consistently avoids being put in any situation where own performance will be inferior to that of the others.

113. Expresses impulses by specific *verbal* "acting out" (e.g., scolding, yelling, cursing, etc.).

114. Exhibits manneristic behavior (tapping on table, biting lips, biting nails, wringing hands, etc.).

115. Is afraid of emotional involvement with others.

116. Has a wish (conscious or unconscious) to kill people who thwart self in any way.

117. Has a wish (conscious or unconscious) to take others' possessions from them.

118. Undervalues and consistently derogates the opposite sex.

119. Psychic conflicts are represented in somatic symptoms.

120. Handles anxieties and conflicts by refusing to recognize their presence.

121. Possesses a basic insecurity and need for attention. Search for "love" is a compulsive or neurotic search for security.

122. Overreacts to danger or makes emergency responses in the absence of actual danger.

123. Is made anxious or disturbed by impulses to commit a criminal act or hostile act (e.g., desire to stab, beat, or kill someone, to set a fire, to mutilate an animal).

124. Fears of phobias present (include all the common fears such as claustrophobia, school phobia, etc.). Continuum ranging from slight anxiety to severe inhibition of activity because of the fears.

125. Has superior intellectual ability (based on clinical observations of functioning level *only*).

126. Repressive mechanism functions adequately.

127. Easy to talk to and get along with in this kind of relationship.

128. Has the capacity for forming close interpersonal relationships.

129. Has an exaggerated need for affection.

130. Ego strength (continuum ranging from severe ego weakness through moderate strength to exceptionally strong ego development).

131. Has obsessional character problems.

132. Has developed defenses which themselves cause suffering.

133. Is a reliable informant.

134. Is provocative.

135. Is "normal," healthy, symptom free (psychologically).